GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 13, 2004, 08:14:32 ; Search time 20 Seconds
(without alignments)
817.628 Million cell updates/sec

Title: US-09-171-607A-1
Perfect score: 893
Sequence: 1 VALNSPLSGGMRGIRGADFQ.....ASCHHAYIVLCIENSFMTAS 170

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	893	100.0	684	2	A53019	collagen alpha 1(X
2	778	87.1	1774	2	B56101	collagen alpha 1(X
3	775	86.8	1315	2	A56101	collagen alpha 1(X
4	504	56.4	1388	2	A53317	collagen alpha 1(X
5	357	40.0	650	2	T22002	hypothetical prote
6	79	8.8	477	2	T05202	pectinesterase hom
7	77	8.6	314	2	AG3054	succinoglycan bios
8	77	8.6	314	2	F98231	succinoglycan bios
9	76	8.5	204	2	B45022	CRK-I - human
10	76	8.5	636	2	JW0047	class I cytokinase
11	75.5	8.5	904	2	T04377	probable pullulana
12	75	8.4	351	2	B89781	conserved hypothet
13	74.5	8.3	850	2	S56015	gastric mucin MUC5
14	73.5	8.2	1653	2	B91052	hypothetical prote
15	73.5	8.2	1653	2	F85896	hypothetical prote
16	73.5	8.2	1653	2	G65028	hypothetical prote
17	73	8.2	435	2	JC5584	hyaluronon glucosam
18	73	8.2	462	2	T17480	endo-xylanase homo
19	73	8.2	513	1	A45333	exopolyphosphatase
20	73	8.2	513	2	D91049	exopolyphosphatase
21	73	8.2	513	2	H85893	exopolyphosphatase
22	72.5	8.1	884	2	C70729	hypothetical prote
23	72	8.1	239	2	A46243	epidermal growth f
24	72	8.1	297	2	AE3128	carbohydrate kinas
25	72	8.1	304	2	A45022	CRK-II - human
26	72	8.1	304	2	I58394	c-Crk - mouse
27	72	8.1	319	2	D98159	hypothetical prote
28	72	8.1	800	2	AG0438	probable Rhs acces
29	72	8.1	1006	2	T00050	hypothetical prote
30	71.5	8.0	554	2	T06374	probable pectinest
31	71.5	8.0	554	2	T06468	pectinesterase (EC
32	71	8.0	579	2	A70954	hypothetical prote
33	71	8.0	644	2	A72519	probable 2-oxoacid
34	71	8.0	732	2	AF0439	probable Rhs acces
35	71	8.0	787	2	G81692	inner membrane pro
36	70.5	7.9	774	2	T04892	probable anthranil
37	70	7.8	234	2	C75368	conserved hypothet
38	70	7.8	375	2	F83906	unsaturated glucur
39	70	7.8	451	2	H75593	hypothetical prote
40	70	7.8	672	2	T36083	hypothetical prote
41	70	7.8	741	2	AC0094	conserved hypothet
42	70	7.8	777	2	AD0982	biotin sulfoxide r
43	69.5	7.8	168	2	S72898	hypothetical prote
44	69.5	7.8	364	2	H87212	probable glucose e
45	69	7.7	266	2	S22511	chlorophyll a/b-bi

ALIGNMENTS

RESULT 1
A53019
collagen alpha 1(XVIII) chain - human (fragment)
N;Contains: endostatin
C;Species: Homo sapiens (man)
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 15-Sep-2003
C;Accession: A53019
R;Oh, S.P.; Warman, M.L.; Seldin, M.F.; Cheng, S.D.; Knoll, J.H.M.; Timmons, S.; Olsen
Genomics 19, 494-499, 1994
A;Title: Cloning of cDNA and genomic DNA encoding human type XVIII collagen and locali
A;Reference number: A53019; MUID:94245237; PMID:8188291
A;Accession: A53019
A;Molecule type: mRNA
A;Residues: 1-684 <OHA>
A;Cross-references: GB:L22548; NID:g348908; PIDN:AAA51864.1; PID:g562794
A;Note: the cited accession number, L25548, is not in Genbank release 103
A;Note: in the authors' translation, 482-Gly is not shown, residues 483-490 are shifte
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
lated and subsequently O-glycosylated.
C;Comment: Different splice forms of collagen alpha 1(XVIII) may be involved in periva
C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of
ay be useful in treating solid tumors.
C;Genetics:
A;Gene: GDB:COL18A1
A;Cross-references: GDB:138752; OMIM:120328
A;Map position: 21q22.3-21q22.3
C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteogl
F;1-684/Product: collagen alpha 1(XVIII) chain (fragment) #status predicted <MAT>
F;1-59/Domain: collagenous (fragment) #status predicted <CO4>
F;74-115/Domain: collagenous #status predicted <CO5>
F;129-201/Domain: collagenous #status predicted <CO6>
F;212-244/Domain: collagenous #status predicted <CO7>
F;257-278/Domain: collagenous #status predicted <CO8>
F;262-264/Region: cell attachment (R-G-D) motif
F;286-340/Domain: collagenous #status predicted <CO9>
F;354-371/Domain: collagenous #status predicted <CO10>
F;502-684/Product: endostatin #status predicted <EST>
F;509-684/Region: multiplexin collagen carboxyl-terminal homologous

Query Match 100.0%; Score 893; DB 2; Length 684;
Best Local Similarity 100.0%; Pred. No. 1e-81;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 573

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        574 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 633

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
```

```
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       634 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 683
```

RESULT 2
B56101
collagen alpha 1(XVIII) chain precursor, long splice form - mouse
N;Contains: collagen alpha 1(XVIII) chain precursor, medium splice form; endostatin
C;Species: Mus musculus (house mouse)
C;Date: 03-Oct-1995 #sequence_revision 08-May-1998 #text_change 15-Sep-2003
C;Accession: B56101; C56101; S72450; S65595; PN0675; A54072; A58816
R;Rehn, M.; Pihlajaniemi, T.
J. Biol. Chem. 270, 4705-4711, 1995
A;Title: Identification of three N-terminal ends of type XVIII collagen chains and tissu
tif homologous to rat and Drosophila frizzled proteins.
A;Reference number: A56101; MUID:95181468; PMID:7876242
A;Accession: B56101
A;Molecule type: mRNA
A;Residues: 1-562 <REH1>
A;Cross-references: GB:U11637; NID:g618429; PIDN:AAC52179.1; PID:g618430
A;Experimental source: splice form clone PE17.24
A;Accession: C56101
A;Molecule type: mRNA
A;Residues: 1-239,487-562 <REH2>
A;Cross-references: GB:U11637; NID:g618429
A;Experimental source: splice form clones PE8.1, PE19, PE15.2
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
submitted to the EMBL Data Library, August 1993
A;Reference number: S72450
A;Accession: S72450
A;Molecule type: mRNA
A;Residues: 487-1146,'L',1148-1193,'F',1195-1210,'R',1212-1512,'L',1514-1522,'F',1524-16
A;Cross-references: EMBL:L22545; NID:g348968; PIDN:AAA19787.1; PID:g511298
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
A;Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa
A;Reference number: A58370; MUID:94240111; PMID:8183893
A;Accession: S65595
A;Molecule type: mRNA
A;Residues: 487-1512,'L',1514-1522,'F',1524-1683,'V',1685-1774 <OH2>
A;Cross-references: EMBL:L22545
R;Abe, N.; Muragaki, Y.; Yoshioka, H.; Inoue, H.; Ninomiya, Y.
Biochem. Biophys. Res. Commun. 196, 576-582, 1993
A;Title: Identification of a novel collagen chain represented by extensive interruptions
A;Reference number: PN0675; MUID:94059075; PMID:8240330
A;Accession: PN0675
A;Molecule type: mRNA
A;Residues: 635-1774 <ABE>
R;Rehn, M.; Hintikka, E.; Pihlajaniemi, T.
J. Biol. Chem. 269, 13929-13935, 1994
A;Title: Primary structure of the alpha1 chain of mouse type XVIII collagen, partial str
collagen chain.
A;Reference number: A54072; MUID:94245707; PMID:8188673
A;Accession: A54072
A;Molecule type: DNA; mRNA
A;Residues: 1293-1403,'R',1405-1774 <REH3>
A;Cross-references: GB:U03714; NID:g487733; PIDN:AAA20657.1; PID:g487734
R;O'Reilly, M.S.; Boehm, T.; Shing, Y.; Fukai, N.; Vasios, G.; Lane, W.S.; Flynn, E.; Bi
Cell 88, 277-285, 1997
A;Title: Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.
A;Reference number: A58816; MUID:97160848; PMID:9008168
A;Accession: A58816
A;Molecule type: protein
A;Residues: 1591-1610 <ORE>
A;Experimental source: hemangioendothelium cells
A;Note: inhibits endothelial cell proliferation
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
lated and subsequently O-glycosylated.
C;Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in peri
C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of un
ay be useful in treating solid tumors.
C;Genetics:
A;Gene: MGI:Col18a1
A;Cross-references: MGI:71175
A;Map position: 10:41.0
A;Introns: 1295/3; 1310/1; 1331/1; 1345/3; 1388/3; 1437/1; 1461/3; 1505/3; 1516/3; 159
A;Note: the list of introns is incomplete
C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteogl
F;1-1774/Product: collagen alpha 1(XVIII) chain precursor, long splice form #status pr
F;1-239,487-1774/Product: collagen alpha 1(XVIII) chain precursor, medium splice form
F;1-24/Domain: signal sequence #status predicted <SIG>
F;786-812/Domain: collagenous #status predicted <CO01>
F;823-896/Domain: collagenous #status predicted <CO02>
F;921-1042/Domain: collagenous #status predicted <CO03>
F;1066-1148/Domain: collagenous #status predicted <CO04>
F;1163-1204/Domain: collagenous #status predicted <CO05>
F;1218-1290/Domain: collagenous #status predicted <CO06>
F;1301-1333/Domain: collagenous #status predicted <CO07>
F;1346-1369/Domain: collagenous #status predicted <CO08>
F;1351-1353/Region: cell attachment (R-G-D) motif
F;1377-1428/Domain: collagenous #status predicted <CO09>
F;1442-1459/Domain: collagenous #status predicted <CO10>
F;1591-1774/Product: endostatin #status predicted <EST>
F;1598-1774/Region: multiplexin collagen carboxyl-terminal homologous
F;354,361,947/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;699,704,1716/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;910,913,1053/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

```
  Query Match             87.1%;  Score 778;  DB 2;  Length 1774;
  Best Local Similarity   85.8%;  Pred. No. 1.2e-69;
  Matches 145;  Conservative   13;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
                ||||:||||||||||||||||||||||||:||||||||||||||||||||||  :||||
Db       1603 VALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRGSVPIV 1662

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
                |||||:| |||::||||:| |:||||||||||:||||||  ||||||||||||:|||  |
Db       1663 NLKDEVLSPSWDSLFSGSQGQLQPGARIFSFDGRDVLRHPAWPQKSVWHGSDPSGRRLME 1722

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 169
              ||||||||    |||||||| ||||  |  |||||::||||||||||||:
Db       1723 SYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMTS 1771
```

RESULT 3
A56101
collagen alpha 1(XVIII) chain precursor, short splice form - mouse
N;Contains: endostatin
C;Species: Mus musculus (house mouse)
C;Date: 03-Oct-1995 #sequence_revision 08-May-1998 #text_change 15-Sep-2003
C;Accession: A56101; A58371; S72450; S65595
R;Rehn, M.; Pihlajaniemi, T.
J. Biol. Chem. 270, 4705-4711, 1995
A;Title: Identification of three N-terminal ends of type XVIII collagen chains and tis
tif homologous to rat and Drosophila frizzled proteins.
A;Reference number: A56101; MUID:95181468; PMID:7876242
A;Accession: A56101
A;Molecule type: mRNA
A;Residues: 1-103 <REH1>
A;Cross-references: GB:U11636; NID:g618427; PIDN:AAC52178.1; PID:g618428
R;Rehn, M.; Pihlajaniemi, T.
Proc. Natl. Acad. Sci. U.S.A. 91, 4234-4238, 1994
A;Title: Alpha1(XVIII), a collagen chain with frequent interruptions in the collagenous
A;Reference number: A58371; MUID:94240112; PMID:8183894
A;Accession: A58371
A;Molecule type: mRNA
A;Residues: 1-928 <REH2>
A;Cross-references: GB:L16898; NID:g404754; PIDN:AAA37434.1; PID:g553894
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
submitted to the EMBL Data Library, August 1993
A;Reference number: S72450
A;Accession: S72450
A;Molecule type: mRNA
A;Residues: 28-687,'L',689-734,'F',736-751,'R',753-1315 <OHW>
A;Cross-references: EMBL:L22545; NID:g348968; PIDN:AAA19787.1; PID:g511298

R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
A;Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa
A;Reference number: A58370; MUID:94240111; PMID:8183893
A;Accession: S65595
A;Molecule type: mRNA
A;Residues: 28-1315 <OHS>
A;Cross-references: EMBL:L22545
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
lated and subsequently O-glycosylated.
C;Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in peri
C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of un
ay be useful in treating solid tumors.
C;Genetics:
A;Gene: MGI:Col18a1
A;Cross-references: MGI:71175
A;Map position: 10:41.0
C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc
F;1-25/Domain: signal sequence #status predicted <SIG>
F;24-235/Region: thrombospondin amino-terminal homologous
F;26-1315/Product: collagen alpha 1(XVIII) chain, short splice form #status predicted <M
F;327-353/Domain: collagenous #status predicted <CO1>
F;364-437/Domain: collagenous #status predicted <CO2>
F;462-583/Domain: collagenous #status predicted <CO3>
F;607-689/Domain: collagenous #status predicted <CO4>
F;704-745/Domain: collagenous #status predicted <CO5>
F;759-831/Domain: collagenous #status predicted <CO6>
F;842-874/Domain: collagenous #status predicted <CO7>
F;887-910/Domain: collagenous #status predicted <CO8>
F;892-894/Region: cell attachment (R-G-D) motif
F;918-969/Domain: collagenous #status predicted <CO9>
F;983-1000/Domain: collagenous #status predicted <CO10>
F;1132-1315/Product: endostatin #status predicted <EST>
F;1139-1315/Region: multiplexin collagen carboxyl-terminal homologous
F;126,488/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;172-228/Disulfide bonds: #status predicted
F;240,245,1257/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;451,454,594/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 86.8%; Score 775; DB 2; Length 1315;
Best Local Similarity 85.2%; Pred. No. 1.6e-69;
Matches 144; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

```
Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||:||||||||||||||||||||||:||||||||||||||||||||| :||||
Db       1144 VALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRGSVPIV 1203

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              |||||:| |||::|||||:| ::||||||||||:||||| ||||||||||||:|||| |
Db       1204 NLKDEVLSPSWDSLFSGSQGQVQPGARIFSFDGRDVLRHPAWPQKSVWHGSDPSGRRLME 1263

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 169
              |||||||||   |||||||| |||| | |||||::||||||||||||:
Db       1264 SYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMTS 1312
```

RESULT 4
A53317
collagen alpha 1(XV) chain precursor - human
N;Alternate names: procollagen alpha 1(XV) chain
C;Species: Homo sapiens (man)
C;Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 15-Sep-2003
C;Accession: A53317; A53146; S28778
R;Kivirikko, S.; Heinaemaeki, P.; Rehn, M.; Honkanen, N.; Myers, J.C.; Pihlajaniemi, T.
J. Biol. Chem. 269, 4773-4779, 1994
A;Title: Primary structure of the alpha1 chain of human type XV collagen and exon-intron
A;Reference number: A53317; MUID:94148920; PMID:8106446
A;Accession: A53317
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1388 <KIV>
A;Cross-references: GB:L25280
A;Note: nucleotide sequence and conceptual translation not complete

R;Muragaki, Y.; Abe, N.; Ninomiya, Y.; Olsen, B.R.; Ooshima, A.
J. Biol. Chem. 269, 4042-4046, 1994
A;Title: The human alpha1(XV) collagen chain contains a large amino-terminal non-triple
A;Reference number: A53146; MUID:94140817; PMID:8307960
A;Accession: A53146
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-9,'S',11-48,'V',50-94,'A',96-149,'A',151-203,'V',205-408,'A',410-569 <MUI
A;Cross-references: GB:D21230; NID:g415605; PIDN:BAA04762.1; PID:d1005294; PID:g460703
R;Myers, J.C.; Kivirikko, S.; Gordon, M.K.; Dion, A.S.; Pihlajaniemi, T.
Proc. Natl. Acad. Sci. U.S.A. 89, 10144-10148, 1992
A;Title: Identification of a previously unknown human collagen chain, alpha1(XV), chara
A;Reference number: S28778; MUID:93066196; PMID:1279671
A;Accession: S28778
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 544-640,'P',642-811,'P',813-1252 <MYE>
C;Genetics:
A;Gene: GDB:COL15A1
A;Cross-references: GDB:132578; OMIM:120325
A;Map position: 9q21-9q22
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-1388/Product: collagen alpha 1(XV) chain #status predicted <MAT>
F;1216-1388/Region: multiplexin collagen carboxyl-terminal homologous

Query Match 56.4%; Score 504; DB 2; Length 1388;
Best Local Similarity 56.9%; Pred. No. 3.3e-42;
Matches 95; Conservative 27; Mismatches 41; Indels 4; Gaps 1;

```
Qy          2 ALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIVN 61
              ||| | || :|    ||||||:|||| ||  |:|||||| |||| :|||:|:| ::||||
Db       1222 ALNMPFSGDIR----ADFQCFKQARAAGLLSTYRAFLSSHLQDLSTIVRKAERYSLPIVN 1277

Qy         62 LKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTES 121
              || ::|| :|:::||| |        |:||||:|::  |:|||| :|||| |:| || ::
Db       1278 LKGQVLFNNWDSIFSGHGGQFNMHIPIYSFDGRDIMTDPSWPQKVIWHGSSPHGVRLVDN 1337

Qy        122 YCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 168
              ||| |||   : || || |  |::| | | || :  ||||||||||
Db       1338 YCEAWRTADTAVTGLASPLSTGKILDQKAYSCANRLIVLCIENSFMT 1384
```

RESULT 5
T22002
hypothetical protein F39H11.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T22002
R;White, S.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19500
A;Accession: T22002
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-650 <WIL>
A;Cross-references: EMBL:Z81079; PIDN:CAB03084.1; GSPDB:GN00019; CESP:F39H11.4
A;Experimental source: clone F39H11
C;Genetics:
A;Gene: CESP:F39H11.4
A;Map position: 1
A;Introns: 109/3; 154/1; 357/1; 420/3; 464/3; 566/2; 594/1; 628/3

Query Match 40.0%; Score 357; DB 2; Length 650;
Best Local Similarity 44.0%; Pred. No. 8.1e-28;
Matches 73; Conservative 24; Mismatches 63; Indels 6; Gaps 4;

```
Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              :||: | || : |:|||| ||:::||| |    |||| ||| :|||  ||    |     :|
Db        468 IALSQPFSGNLHGLRGADLQCYREARAAGYTTTFRAMLSSNVQDLVRIVHSVD-FDTTVV 526

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              |:    ||||| :  :|::  : | |::|||| |||    || | |||||   | |  |
Db        527 NVAGHHLFPSWRSFVNGAQ--MNPHAKLFSFDRHDVLNDSRWPDKRVWHGSKDGGIR-AE 583
```

```
Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAAS--CHHAYIVLCIEN 164
              ||: ||    | |  |  :     : ||: |  | :  :|||:||
Db        584 QYCDGWRRADSSLTSLAGHISSNTSIFQSSGSEKCENKLVVLCVEN 629
```

RESULT 6
T05202
pectinesterase homolog F4I10.150 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 24-Nov-1999
C;Accession: T05202
R;Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Hoheisel, J.; Mew
submitted to the Protein Sequence Database, February 1999
A;Reference number: Z15402
A;Accession: T05202
A;Molecule type: DNA
A;Residues: 1-477 <BEV>
A;Cross-references: EMBL:AL035525
A;Experimental source: cultivar Columbia; BAC clone F4I10
C;Genetics:
A;Map position: 4
A;Introns: 7/3; 47/1; 90/1; 246/1; 354/1
A;Note: F4I10.150
C;Superfamily: pectinesterase

  Query Match 8.8%; Score 79; DB 2; Length 477;
  Best Local Similarity 21.9%; Pred. No. 5.2;
  Matches 32; Conservative 25; Mismatches 53; Indels 36; Gaps 6;

```
Qy          3 LNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIVNL 62
              |:: ||     :| | |          | :|  :: ::  |  |||::|      :|:|
Db         63 LSAALSNQATCMEGFD----------GTSGLVKSLVAGSLDQLYSMLRE----LLPLVQ- 107

Qy         63 KDELLFPSWEALFSGSEGPL------KPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGR 116
                 |  :       ||:       || ::   |  : |: | | :         |: |
Db        108 ------PEQKPKAVSKPGPIAKGPKAPPGRKLRDTDEDESLQFPDWVR--------PDDR 153

Qy        117 RLTESYCETWRTE-APSATGQASSLL 141
              :| ||   |:    |   ||  : ::
Db        154 KLLESNGRTYDVSVALDGTGNFTKIM 179
```

RESULT 7
AG3054
succinoglycan biosynthesis protein exoM [imported] - Agrobacterium tumefaciens (strain C
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AG3054
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AG3054
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-314 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL44853.1; PID:g17742499; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: exoM
A;Map position: linear chromosome
C;Superfamily: Rhizobium succinoglycan biosynthesis glycosyltransferase

  Query Match 8.6%; Score 77; DB 2; Length 314;
  Best Local Similarity 27.9%; Pred. No. 5;
  Matches 29; Conservative 14; Mismatches 41; Indels 20; Gaps 5;

```
Qy         64 DELLFPSW------EALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDP---N 114
              ||   | |       |  :|:|  | |   ::  :       | | ::  :| : |   |
Db        105 DETAPPHWLTALLETAEETGAETVLGPVTAVYRDNA------PGWMKRGDFHSTVPVWVN 158

Qy        115 GRRLTESYCET-WRTEAPSATGQASSLLGGRLLGQSAASCHHAY 157
              |  :|   | |  | |||:  |:  :|     ||||      | :
Db        159 GEIITGYTCNTLLRMEAPAVKGRRFAL----ALGQSGGEDTHFF 198
```

RESULT 8
F98231
succinoglycan biosynthesis protein exoM [imported] - Agrobacterium tumefaciens (strain
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C;Accession: F98231
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldmar
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium t
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: F98231
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-314 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK89376.1; PID:g15159227; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_1606
A;Map position: linear chromosome
C;Superfamily: Rhizobium succinoglycan biosynthesis glycosyltransferase

  Query Match 8.6%; Score 77; DB 2; Length 314;
  Best Local Similarity 27.9%; Pred. No. 5;
  Matches 29; Conservative 14; Mismatches 41; Indels 20; Gaps 5;

```
Qy         64 DELLFPSW------EALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDP---N 114
              ||   | |       |  :|:|  | |   ::  :       | | ::  :| : |   |
Db        105 DETAPPHWLTALLETAEETGAETVLGPVTAVYRDNA------PGWMKRGDFHSTVPVWVN 158

Qy        115 GRRLTESYCET-WRTEAPSATGQASSLLGGRLLGQSAASCHHAY 157
              |  :|   | |  | |||:  |:  :|     ||||      | :
Db        159 GEIITGYTCNTLLRMEAPAVKGRRFAL----ALGQSGGEDTHFF 198
```

RESULT 9
B45022
CRK-I - human
C;Species: Homo sapiens (man)
C;Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 12-Feb-1999
C;Accession: B45022
R;Matsuda, M.; Tanaka, S.; Nagata, S.; Kojima, A.; Kurata, T.; Shibuya, M.
Mol. Cell. Biol. 12, 3482-3489, 1992
A;Title: Two species of human CRK cDNA encode proteins with distinct biological activi
A;Reference number: A45022; MUID:92334347; PMID:1630456
A;Accession: B45022
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-204 <MAT>
A;Experimental source: embryonic lung cells
A;Note: sequence extracted from NCBI backbone (NCBIN:108771, NCBIP:108772)
C;Superfamily: crk transforming protein; SH2 homology; SH3 homology
F;13-118/Domain: SH2 homology <SH2>
F;139-187/Domain: SH3 homology <SH3>

  Query Match 8.5%; Score 76; DB 2; Length 204;
  Best Local Similarity 29.0%; Pred. No. 3.8;
  Matches 20; Conservative 12; Mismatches 21; Indels 16; Gaps 3;

```
Qy         88 IFSFDGK----------DVLRHPTWPQKSVWHGSDPNGRR--LTESYCETWRTEAPSATG 135
              :| |:|           |:||    |::  |:  |  |:|  :   | | :|     |:
Db        140 LFDFNGNDEEDLPFKKGDILRIRDKPEEQWWNAEDSEGKRGMIPVPYVEKYR----PASA 195

Qy        136 QASSLLGGR 144
                |:|:|||
Db        196 SVSALIGGR 204
```

```
RESULT 10
JW0047
class I cytokinase receptor precursor - human
N;Alternate names: WSX-1
C;Species: Homo sapiens (man)
C;Date: 17-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 21-Jul-2000
C;Accession: JW0047
R;Sprecher, C.A.; Grant, F.J.; Baumgartner, J.W.; Presnell, S.R.; Schrader, S.K.; Yamagi
Biochem. Biophys. Res. Commun. 246, 82-90, 1998
A;Title: Cloning and characterization of a novel class I cytokine receptor.
A;Reference number: JW0047; MUID:98262921; PMID:9600072
A;Accession: JW0047
A;Molecule type: mRNA
A;Residues: 1-636 <SPR>
A;Cross-references: GB:AF053004; NID:g3153240; PIDN:AAC39755.1; PID:g3153241
A;Experimental source: brain
C;Genetics:
A;Map position: 19p13.11
C;Keywords: glycoprotein
F;1-32/Domain: signal sequence #status predicted <SIG>
F;515-540/Domain: transmembrane #status predicted <TMM>
F;554-561/Domain: cytoplasmic #status predicted <CTP>
F;51,76,302,311,374,382,467/Binding site: carbohydrate (Asn) (covalent) #status predicte

  Query Match             8.5%;  Score 76;  DB 2;  Length 636;
  Best Local Similarity   24.2%;  Pred. No. 15;
  Matches   39;  Conservative   12;  Mismatches   46;  Indels   64;  Gaps    7;

Qy         11 MRGIRGADF----------------QCFQQARAVGLAGTFRAFLSSRLQDL---------- 45
              ||| ||| |                ||: |  | ||  : :    | ||
Db          1 MRGGRGAPFWLWPLPKLALLPLLWVLFQRTRPQGSAGPLQCYGVGPLGDLNCSWEPLGDL 60

Qy         46 ------------------YSIVRRADRAAVPI----VNLKDELLF-------PSWEALFS 76
                              ::   | |: | |    : : |:||        | |  :|
Db         61 GAPSELHLQSQKYRSNKTQTVAVAAGRSWVAIPREQLTMSDKLLVWGTKAGQPLWPPVFV 120

Qy         77 GSEGPLKPGA-RI---FSFDGKDVLR------HPTWPQKSV 107
                |  :|| | |:     |   | |         ||||   |
Db        121 NLETQMKPNAPRLGPDVDFSEDDPLEATVHWAPPTWPSHKV 161


RESULT 11
T04377
probable pullulanase (EC 3.2.1.41) - barley
N;Alternate names: pullulanase
C;Species: Hordeum vulgare (barley)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 22-Jun-2003
C;Accession: T04377
R;Lok, F.; Kristensen, M.; Planchot, V.; Leah, R.; Svendsen, I.; Svenson, B.
submitted to the EMBL Data Library, December 1997
A;Description: Isolation and characterization of starch debranching enzyme, limit dextri
A;Reference number: Z15320
A;Accession: T04377
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-904 <LOK>
A;Cross-references: EMBL:AF022725; NID:g2502057; PIDN:AAD04189.1; PID:g2677837
A;Experimental source: cv. Igri
C;Genetics:
A;Gene: HvLD99
A;Introns: 21/3; 70/1; 87/3; 124/1; 169/3; 235/2; 285/2; 305/3; 346/1; 376/2; 407/3; 439
C;Superfamily: pullulanase type debranching enzyme
C;Keywords: glycosidase; hydrolase

  Query Match             8.5%;  Score 75.5;  DB 2;  Length 904;
  Best Local Similarity   22.5%;  Pred. No. 25;
  Matches   45;  Conservative   19;  Mismatches   67;  Indels   69;  Gaps    8;

Qy          5 SPLSGGMRGIRGADFQCFQQARAVGLAGT----------FRAFLSSRLQDLYSIVR---- 50
              || :|   ||:| | :   |  : ||  |          :|||       |: |:|:
Db         62 SPSNG---GIQGYDSKVELQPESAGLPETVTQKFPFISSYRAFKVPSSVDVASLVKCQLV 118

Qy         51 ----RADRAAVPIVNL--------------------KDELLFPSWE--------ALFSG 77
                  ||   | :  |                      :| :    |          | | |
Db        119 VASFGADGKHVDVTGLQLPGVLDDMFAYTGPLGAVFSEDSVSLHLWAPTAQGVSVCFFDG 178

Qy         78 SEGP------LKPGARIFSFDGK------------DVLRHPTWPQKSVWHGSDPNGRRLT 119
                ||      ||    ::|  |             ||  |||  |       ||  | |:
Db        179 PAGPALETVQLKESNGVWSVTGPREWENRYYLYEVDVY-HPTKAQVLKCLAGDPYARSLS 237

Qy        120 ESYCETWRTEAPSATGQASS 139
                :   ||  :  : | : :|
Db        238 ANGARTWLVDINNETLKPAS 257


RESULT 12
B89781
conserved hypothetical protein SA0184 [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: B89781
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogu
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Stapylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: B89781
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-351 <KUR>
A;Cross-references: GB:BA000018; PID:g13700106; PIDN:BAB41405.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: SA0184

  Query Match             8.4%;  Score 75;  DB 2;  Length 351;
  Best Local Similarity   23.3%;  Pred. No. 9.2;
  Matches   42;  Conservative   17;  Mismatches   67;  Indels   54;  Gaps    8;

Qy         11 MRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDL---------YSIVRRADRA-AVPIV 60
              :  |      :|   |  :      |  |:|  | |          ::   | |   :| :|
Db         97 IEAIMAQGLKCCLNASIIS-----RELLTSLHQQLNDFTLLSFCHNYYPRPDTGLSVDLV 151

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              | |:||::               | |:|: |      || |         |   |
Db        152 NKKNELIY------------QFNPKAQIYGFIVGSGLRGPL-------HKGLPT------ 186

Qy        121 SYCETWRTEAP-------SATGQASSLLGGRLLGQSAAS-----CHHAYIVLCIENSFMT 168
              |  |  |           || :  |:|  |:       |       | | :  ||||  | |
Db        187 --IEATRHSHPVVAAKLLQETGVSEVLVGDSLIEMRQAKQLIDFCKHRHFTLCIEEVFDT 244


RESULT 13
S56015
gastric mucin MUC5AC - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Apr-2000
C;Accession: S56015; S53361
R;Klomp, L.W.J.; van Rens, L.; Strous, G.J.
Biochem. J. 308, 831-838, 1995
A;Title: Cloning and analysis of human gastric mucin cDNA reveals two types of conserve
A;Reference number: S56015; MUID:97104281; PMID:8948439
A;Accession: S56015
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-850 <KLO>
A;Cross-references: EMBL:X81649; NID:g547516; PIDN:CAA57309.1; PID:g547517
R;Guyonnet-Duperat, V.; Audie, J.P.; Debailleul, V.; Laine, A.; Buisine, M.P.; Galiegue
Biochem. J. 305, 211-219, 1995
A;Title: Characterization of the human mucin gene MUC5AC: a consensus cysteine-rich dom
A;Reference number: S53361; MUID:95126907; PMID:7826332
A;Accession: S53361
A;Status: preliminary; nucleic acid sequence not shown
```

A;Molecule type: mRNA
A;Residues: 648-678,'L',680-733,'L',735-760 <GUY>
A;Cross-references: EMBL:Z34280; NID:g563380; PIDN:CAA84034.1; PID:g563381
A;Experimental source: clone JUL32
A;Note: this publication is not cited in GenBank entry HSMUCIN5, release 113.0

Query Match 8.3%; Score 74.5; DB 2; Length 850;
Best Local Similarity 24.7%; Pred. No. 30;
Matches 39; Conservative 23; Mismatches 71; Indels 25; Gaps 6;

```
Qy          13 GIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADR-AAVPIVNLKDELLFPSW 71
               || | ||  ||  |  |   ||        :   |:  ||:     |: :|  :::
Db         592 GINGGDFDTFQNLRDEGY--TF-------CESPRSVQCRAESFPNTPLADLGQDVICSHT 642

Qy          72 EALFSGSEGPLKP-----GARIFSFDGKDVLRHPTWPQKSVW------HGSDPNGRRLTE 120
               | |   ::  | |       ||   :  :| |  | | |:|       | :   :
Db         643 EGLICLNKNQLPPICYNYEIRIQCCETVNVCRDITRPPKTVATTRPTPHPTGAQTQTTFT 702

Qy         121 SYCETWRTEAPSATGQ----ASSLLGGRLLGQSAASCH 154
               ::  :  || |:|| :    |:|:  |        :||
Db         703 THMPSASTEQPTATSRGGPTATSVTQGTHTTPVTRNCH 740
```

RESULT 14
B91052
hypothetical protein ECs3386 [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: B91052
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: B91052
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1653 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB36809.1; PID:g13362856; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs3386

Query Match 8.2%; Score 73.5; DB 2; Length 1653;
Best Local Similarity 24.4%; Pred. No. 84;
Matches 32; Conservative 21; Mismatches 47; Indels 31; Gaps 5;

```
Qy          29 GLAGTFRAFLSSRLQDLY---------SIVRRADRAAVPIVNLKDELLFPSWEALFSGSE 79
               |  |  : |     :|||          ::| ||  |:|   :| ::: |  : | |
Db         380 GAPGYSKQFFMFGPRDLYRPGETVILNGLLRDADGKALPNQPIKLDVIKPDGQVLRSVVS 439

Qy          80 GPLKPGARIFSFDGKDVLRHPTWPQKS-----VWH---GSDPNGRRLTESYCETWRTE-- 129
                |            :: | | |||  |     :||    :  |  |: : : | :  |
Db         440 QP------------ENGLYHFTWPLDSNAATGMWHIRANTGDNQYRMWDFHVEDFMPERM 487

Qy         130 APSATGQASSL 140
               | : ||: : |
Db         488 ALNLTGEKTPL 498
```

RESULT 15
F85896
hypothetical protein Z3787 [imported] - Escherichia coli (strain O157:H7, substrain EDL9
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: F85896
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: F85896
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1653 <STO>
A;Cross-references: GB:AE005174; NID:g12516921; PIDN:AAG57634.1; GSPDB:GN00145; UWGP:Z
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z3787

Query Match 8.2%; Score 73.5; DB 2; Length 1653;
Best Local Similarity 24.4%; Pred. No. 84;
Matches 32; Conservative 21; Mismatches 47; Indels 31; Gaps 5;

```
Qy          29 GLAGTFRAFLSSRLQDLY---------SIVRRADRAAVPIVNLKDELLFPSWEALFSGSE 79
               |  |  : |     :|||          ::| ||  |:|   :| ::: |  : | |
Db         380 GAPGYSKQFFMFGPRDLYRPGETVILNGLLRDADGKALPNQPIKLDVIKPDGQVLRSVVS 439

Qy          80 GPLKPGARIFSFDGKDVLRHPTWPQKS-----VWH---GSDPNGRRLTESYCETWRTE-- 129
                |            :: | | |||  |     :||    :  |  |: : : | :  |
Db         440 QP------------ENGLYHFTWPLDSNAATGMWHIRANTGDNQYRMWDFHVEDFMPERM 487

Qy         130 APSATGQASSL 140
               | : ||: : |
Db         488 ALNLTGEKTPL 498
```

Search completed: March 13, 2004, 08:18:27
Job time : 22 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 13, 2004, 08:10:51 ; Search time 18 Seconds
(without alignments)
491.774 Million cell updates/sec

Title: US-09-171-607A-1
Perfect score: 893
Sequence: 1 VALNSPLSGGMRGIRGADFQ.....ASCHHAYIVLCIENSFMTAS 170

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	893	100.0	1516	1	CA1H_HUMAN	P39060 homo sapien
2	778	87.1	1774	1	CA1H_MOUSE	P39061 mus musculu
3	504	56.4	1388	1	CA1E_HUMAN	P39059 homo sapien
4	76	8.5	556	1	GLI_CHICK	P55878 gallus gall
5	75.5	8.5	334	1	XA5L_MOUSE	Q9wtj8 mus musculu
6	75	8.4	271	1	Y4BG_RHISN	P55374 rhizobium s
7	74	8.3	296	1	CRK_XENLA	P87378 xenopus lae
8	74	8.3	780	1	STRN_RAT	P70483 rattus norv
9	73.5	8.2	325	1	XA5L_HUMAN	Q9y247 homo sapien
10	73.5	8.2	1653	1	YFHM_ECOLI	P76578 escherichia
11	73	8.2	512	1	PPX_ECOLI	P29014 escherichia
12	72.5	8.1	884	1	YP67_MYCTU	Q50654 mycobacteri
13	72.5	8.1	884	1	YP97_MYCBO	P59974 mycobacteri
14	72	8.1	304	1	CRK_HUMAN	P46108 homo sapien
15	72	8.1	304	1	CRK_MOUSE	Q64010 mus musculu
16	72	8.1	304	1	CRK_RAT	Q63768 rattus norv
17	72	8.1	613	1	HS75_CANAL	P87222 candida alb
18	71	8.0	787	1	OXAA_CHLMU	Q9pke3 chlamydia m
19	70	7.8	780	1	STRN_MOUSE	O55106 mus musculu
20	70	7.8	837	1	LZTR_MOUSE	Q9cq33 mus musculu
21	70	7.8	953	1	CAR4_MOUSE	Q8bhb0 mus musculu
22	69	7.7	266	1	CB21_SINAL	P13851 sinapis alb
23	69	7.7	309	1	YHCC_ECOLI	P45476 escherichia
24	69	7.7	326	1	TMOE_PSEME	Q00460 pseudomonas
25	69	7.7	462	1	MYCN_MOUSE	P03966 mus musculu
26	69	7.7	494	1	ALG8_PSEAE	Q52463 pseudomonas
27	69	7.7	1233	1	NME3_HUMAN	Q14957 homo sapien
28	68.5	7.7	390	1	PGK_BUCAI	P57525 buchnera ap
29	68.5	7.7	1289	1	C5AB_BACUD	Q45753 bacillus th
30	68	7.6	610	1	MET7_SCHPO	O74314 schizosacch
31	68	7.6	953	1	CAR4_HUMAN	Q9y239 homo sapien
32	68	7.6	999	1	MERK_HUMAN	Q12866 homo sapien
33	67.5	7.6	348	1	NU2M_BRARE	Q9miy9 brachydanio
34	67.5	7.6	536	1	CAR9_HUMAN	Q9h257 homo sapien
35	67.5	7.6	579	1	FAAH_MOUSE	O08914 mus musculu
36	67.5	7.6	692	1	GYRB_BARBA	P94281 bartonella
37	67	7.5	145	1	ANF_RANCA	P18909 rana catesb
38	67	7.5	250	1	HXB9_HUMAN	P17482 homo sapien
39	67	7.5	364	1	YM28_MYCTU	Q10512 mycobacteri
40	67	7.5	577	1	BAG3_MOUSE	Q9jlv1 mus musculu
41	66.5	7.4	185	1	NPW_RAT	Q8klm5 rattus norv
42	66.5	7.4	368	1	GALT_HUMAN	O60755 homo sapien
43	66.5	7.4	449	1	DHE2_CLOSY	P24295 clostridium
44	66.5	7.4	1548	1	SMCY_MOUSE	Q62240 mus musculu
45	66.5	7.4	3670	1	CSM3_HUMAN	Q7z407 homo sapien

ALIGNMENTS

RESULT 1
CA1H_HUMAN
ID CA1H_HUMAN STANDARD; PRT; 1516 AA.
AC P39060; Q9UK38; Q9Y6Q7; Q9Y6Q8;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 1(XVIII) chain precursor [Contains: Endostatin].
GN COL18A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98164096; PubMed=9503365;
RA Saarela J., Ylikarppa R., Rehn M., Purmonen S., Pihlajaniemi T.;
RT "Complete primary structure of two variant forms of human type XVIII
RT collagen and tissue-specific differences in the expression of the
RT corresponding transcripts.";
RL Matrix Biol. 16:319-328(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [3]
RP SEQUENCE OF 834-1516 FROM N.A.
RX MEDLINE=94245237; PubMed=8188291;
RA Oh S.P., Warman M.L., Seldin M.F., Cheng S., Knoll J.H., Timmons S.,
RA Olsen B.R.;
RT "Cloning of cDNA and genomic DNA encoding human type XVIII collagen
RT and localization of the alpha 1(XVIII) collagen gene to mouse
RT chromosome 10 and human chromosome 21.";
RL Genomics 19:494-499(1994).
RN [4]
RP SEQUENCE OF 1334-1516 FROM N.A.
RC TISSUE=Placenta;
RA Zhi-Yong H., Biao L., Wei-Jie Z., Xiang-Fu W.;
RT "Cloning and expression of human endostatin gene in Escherichia
RT coli.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN [5]
RP INVOLVEMENT IN KNOBLOCH SYNDROME.

```
RX   MEDLINE=20400145; PubMed=10942434;
RA   Sertie A.L., Sossi V., Camargo A.A., Zatz M., Brahe C.,
RA   Passos-Bueno M.R.;
RT   "Collagen XVIII, containing an endogenous inhibitor of angiogenesis
RT   and tumor growth, plays a critical role in the maintenance of retinal
RT   structure and in neural tube closure.";
RL   Hum. Mol. Genet. 9:2051-2058(2000).
RN   [6]
RP   VARIANT ASN-1437.
RX   MEDLINE=21518361; PubMed=11606364;
RA   Tughetti P., Suzuki O., Godoi P.H., Alves V.A., Sertie A.L.,
RA   Zorick T., Soares F., Camargo A., Moreira E.S., di Loreto C.,
RA   Moreira-Filho C.A., Simpson A., Oliva G., Passos-Bueno M.R.;
RT   "A polymorphism in endostatin, an angiogenesis inhibitor, predisposes
RT   for the development of prostatic adenocarcinoma.";
RL   Cancer Res. 61:7375-7378(2001).
CC   -!- FUNCTION: COL18A1A probably plays a major role in determining the
CC       retinal structure as well as in the closure of the neural tube.
CC   -!- FUNCTION: Endostatin potently inhibits endothelial cell
CC       proliferation and angiogenesis. May inhibit angiogenesis by
CC       binding to the heparan sulphate proteoglycans involved in growth
CC       factor signalling.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Long; Synonyms=NC-493;
CC         IsoId=P39060-1; Sequence=Displayed;
CC       Name=Short; Synonyms=NC1-303;
CC         IsoId=P39060-2; Sequence=VSP_001155, VSP_001156;
CC   -!- TISSUE SPECIFICITY: Present in multiple organs with highest levels
CC       in liver, lung and kidney.
CC   -!- PTM: Prolines at the third position of the tripeptide repeating
CC       unit (G-X-Y) are hydroxylated in some or all of the chains.
CC   -!- POLYMORPHISM: There is an association between a polymorphism in
CC       position 1437 and prostate cancer. Heterozygous Asn-1437
CC       individuals have a 2.5 times increased chance of developing
CC       prostate cancer as compared with homozygous Asp-1437 individuals.
CC   -!- DISEASE: Defects in COL18A1 are a cause of Knobloch syndrome (KNO)
CC       [MIM:267750]; an autosomal recessive disorder defined by the
CC       occurrence of high myopia, vitreoretinal degeneration with retinal
CC       detachment, macular abnormalities and occipital encephalocele.
CC   -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC       INTERRUPTED HELICES (FACIT) FAMILY.
CC   -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF018081; AAC39658.1; -.
DR   EMBL; AF018082; AAC39659.1; -.
DR   EMBL; AL163302; CAB90482.1; -.
DR   EMBL; L22548; AAA51864.1; -.
DR   EMBL; AF184060; AAF01310.1; ALT_INIT.
DR   PDB; 1BNL; 02-DEC-98.
DR   GlycoSuiteDB; P39060; -.
DR   Genew; HGNC:2195; COL18A1.
DR   MIM; 120328; -.
DR   MIM; 267750; -.
DR   GO; GO:0005581; C:collagen; TAS.
DR   GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR   GO; GO:0008285; P:negative regulation of cell proliferation; TAS.
DR   GO; GO:0007601; P:vision; TAS.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 7.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;
KW   Polymorphism; 3D-structure.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24   1516       COLLAGEN ALPHA 1(XVIII) CHAIN.
FT   CHAIN      1334   1516       ENDOSTATIN.
FT   DOMAIN      221    409       TSP N-TERMINAL.
FT   DOMAIN      410    516       NONHELICAL REGION 1 (NC1).
FT   DOMAIN      517    550       TRIPLE-HELICAL REGION 1 (COL1).
FT   DOMAIN      551    560       NONHELICAL REGION 2 (NC2).
FT   DOMAIN      561    640       TRIPLE-HELICAL REGION 2 (COL2).
FT   DOMAIN      641    664       NONHELICAL REGION 3 (NC3).
FT   DOMAIN      665    786       TRIPLE-HELICAL REGION 3 (COL3).
FT   DOMAIN      787    809       NONHELICAL REGION 4 (NC4).
FT   DOMAIN      810    892       TRIPLE-HELICAL REGION 4 (COL4).
FT   DOMAIN      893    906       NONHELICAL REGION 5 (NC5).
FT   DOMAIN      907    948       TRIPLE-HELICAL REGION 5 (COL5).
FT   DOMAIN      949    961       NONHELICAL REGION 6 (NC6).
FT   DOMAIN      962   1034       TRIPLE-HELICAL REGION 6 (COL6).
FT   DOMAIN     1035   1044       NONHELICAL REGION 7 (NC7).
FT   DOMAIN     1045   1077       TRIPLE-HELICAL REGION 7 (COL7).
FT   DOMAIN     1078   1089       NONHELICAL REGION 8 (NC8).
FT   DOMAIN     1090   1111       TRIPLE-HELICAL REGION 8 (COL8).
FT   DOMAIN     1112   1118       NONHELICAL REGION 9 (NC9).
FT   DOMAIN     1119   1173       TRIPLE-HELICAL REGION 9 (COL9).
FT   DOMAIN     1174   1186       NONHELICAL REGION 10 (NC10).
FT   DOMAIN     1187   1204       TRIPLE-HELICAL REGION 10 (COL10).
FT   DOMAIN     1205   1516       NONHELICAL REGION 11 (NC11).
FT   CARBOHYD     68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    129    129       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    164    164       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    691    691       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1329   1329       O-LINKED (GALNAC. . .).
FT                                /FTId=CAR_000150.
FT   DISULFID   1366   1506       BY SIMILARITY.
FT   DISULFID   1468   1498       BY SIMILARITY.
FT   SITE       1095   1097       CELL ATTACHMENT SITE (POTENTIAL).
FT   VARSPLIC      1    180       Missing (in isoform Short).
FT                                /FTId=VSP_001155.
FT   VARSPLIC    181    215       HTTEAGTLPAPTPSPPSLGRPWAPLTGPSVPPPSS -> MA
FT                                PRCPWPWPRRRRLLDVLAPLVLLLGVRAASAEP (in
FT                                isoform Short).
FT                                /FTId=VSP_001156.
FT   VARIANT    1437   1437       D -> N (increased risk of developing
FT                                prostate cancer).
FT                                /FTId=VAR_012709.
FT   CONFLICT    428    428       F -> S (IN REF. 2).
FT   CONFLICT    841    841       I -> V (IN REF. 2).
FT   CONFLICT    877    877       V -> L (IN REF. 3).
FT   CONFLICT    886    886       P -> R (IN REF. 3).
FT   CONFLICT    912    912       P -> R (IN REF. 3).
FT   CONFLICT    933    933       R -> L (IN REF. 3).
FT   CONFLICT    975    975       P -> L (IN REF. 3).
FT   CONFLICT   1064   1064       A -> P (IN REF. 3).
FT   CONFLICT   1084   1084       L -> K (IN REF. 3).
FT   CONFLICT   1120   1120       P -> A (IN REF. 3).
FT   CONFLICT   1123   1123       P -> A (IN REF. 3).
FT   CONFLICT   1126   1126       P -> PPGP (IN REF. 2).
FT   CONFLICT   1206   1206       G -> GQ (IN REF. 3).
FT   CONFLICT   1304   1304       R -> G (IN REF. 3).
FT   CONFLICT   1314   1314       A -> G (IN REF. 3).
FT   CONFLICT   1323   1324       LR -> CG (IN REF. 3).
FT   CONFLICT   1443   1443       R -> T (IN REF. 4).
FT   CONFLICT   1483   1483       S -> Y (IN REF. 4).
SQ   SEQUENCE   1516 AA;  153840 MW;  3C70F29A4476EE76 CRC64;

 Query Match           100.0%;  Score 893;  DB 1;  Length 1516;
 Best Local Similarity  100.0%;  Pred. No. 1.7e-80;
 Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1346 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 1405

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1406 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 1465

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1466 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 1515


RESULT 2
CA1H_MOUSE
ID   CA1H_MOUSE     STANDARD;      PRT;  1774 AA.
AC   P39061; Q60672; Q61437; Q62001; Q62002; Q9JK63;
DT   01-FEB-1995 (Rel. 31, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Collagen alpha 1(XVIII) chain precursor [Contains: Endostatin].
GN   COL18A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 3).
RC   STRAIN=BALB/c; TISSUE=Liver;
RX   MEDLINE=94245707; PubMed=8188673;
RA   Rehn M.V., Hintikka E., Pihlajaniemi T.;
RT   "Primary structure of the alpha 1 chain of mouse type XVIII collagen,
RT   partial structure of the corresponding gene, and comparison of the
RT   alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen
RT   chain.";
RL   J. Biol. Chem. 269:13929-13935(1994).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RX   MEDLINE=96435922; PubMed=8838808;
RA   Rehn M., Hintikka E., Pihlajaniemi T.;
RT   "Characterization of the mouse gene for the alpha-1 chain of type
RT   XVIII collagen (COL18A1) reveals that the three variant N-terminal
RT   polypeptide forms are transcribed from two widely separated
RT   promoters.";
RL   Genomics 32:436-446(1996).
RN   [3]
RP   SEQUENCE OF 1-1387 FROM N.A. (ISOFORM 3).
RX   MEDLINE=94240112; PubMed=8183894;
RA   Rehn M.V., Pihlajaniemi T.;
RT   "Alpha 1(XVIII), a collagen chain with frequent interruptions in the
RT   collagenous sequence, a distinct tissue distribution, and homology
RT   with type XV collagen.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).
RN   [4]
RP   SEQUENCE OF 487-1774 FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=94240111; PubMed=8183893;
RA   Oh S.P., Kamagata Y., Muragaki Y., Timmons S., Ooshima A., Olsen B.R.;
RT   "Isolation and sequencing of cDNAs for proteins with multiple domains
RT   of Gly-Xaa-Yaa repeats identify a distinct family of collagenous
RT   proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233(1994).
RN   [5]
RP   SEQUENCE OF 1591-1774 FROM N.A.
RX   MEDLINE=21217748; PubMed=11321448;
RA   Jia S., Zhu F., Li H., He F., Xiu R.-J.;
RT   "Anticancer treatment of endostatin gene therapy by targeting tumor
RT   neovasculature in C57/BL mice.";
RL   Clin. Hemorheol. Microcirc. 23:251-257(2000).
RN   [6]
RP   CHARACTERIZATION OF ENDOSTATIN, AND PARTIAL SEQUENCE.
RX   MEDLINE=97160848; PubMed=9008168;
RA   O'Reilly M.S., Boehm T., Shing Y., Fukai N., Vasios G., Lane W.S.,
RA   Flynn E., Birkhead J.R., Olsen B.R., Folkman J.;
RT   "Endostatin: an endogenous inhibitor of angiogenesis and tumor
RT   growth.";
RL   Cell 88:277-285(1997).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF ENDOSTATIN.
RX   MEDLINE=98169382; PubMed=9501087;
RA   Hohenester E., Sasaki T., Olsen B.R., Timpl R.;
RT   "Crystal structure of the angiogenesis inhibitor endostatin at 1.5-A
RT   resolution.";
RL   EMBO J. 17:1656-1664(1998).
CC   -!- FUNCTION: Endostatin potently inhibits endothelial cell
CC       proliferation and angiogenesis. May inhibit angiogenesis by
CC       binding to the heparan sulfate proteoglycans involved in growth
CC       factor signaling.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative promoter;
CC         Comment=2 isoforms, 1 (shown here) and 3, are produced by use
CC         of alternative promoters;
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=1; Synonyms=NC1-764;
CC         IsoId=P39061-3; Sequence=Displayed;
CC       Name=2; Synonyms=Long, NC1-517;
CC         IsoId=P39061-1; Sequence=VSP_008303;
CC         Note=Produced by alternative splicing of isoform 1;
CC       Name=3; Synonyms=Short, NC1-301;
CC         IsoId=P39061-2; Sequence=VSP_001157, VSP_001158;
CC   -!- PTM: Prolines at the third position of the tripeptide repeating
CC       unit (G-X-Y) are hydroxylated in some or all of the chains.
CC   -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC       INTERRUPTED HELICES (FACIT) FAMILY.
CC   -!- SIMILARITY: Contains 1 frizzled (FZ) domain.
CC   -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L16898; AAA37434.1; -.
DR   EMBL; U03714; AAA20657.1; -.
DR   EMBL; U03715; AAC52901.1; -.
DR   EMBL; U34606; AAC52901.1; JOINED.
DR   EMBL; U34608; AAC52901.1; JOINED.
DR   EMBL; U34609; AAC52901.1; JOINED.
DR   EMBL; U34610; AAC52901.1; JOINED.
DR   EMBL; U34611; AAC52901.1; JOINED.
DR   EMBL; U34612; AAC52901.1; JOINED.
DR   EMBL; U34613; AAC52901.1; JOINED.
DR   EMBL; U03716; AAC52901.1; JOINED.
DR   EMBL; U03718; AAC52901.1; JOINED.
DR   EMBL; U03715; AAC52902.1; -.
DR   EMBL; U34607; AAC52902.1; JOINED.
DR   EMBL; U34608; AAC52902.1; JOINED.
DR   EMBL; U34609; AAC52902.1; JOINED.
DR   EMBL; U34610; AAC52902.1; JOINED.
DR   EMBL; U34611; AAC52902.1; JOINED.
DR   EMBL; U34612; AAC52902.1; JOINED.
DR   EMBL; U34613; AAC52902.1; JOINED.
DR   EMBL; U03716; AAC52902.1; JOINED.
DR   EMBL; U03718; AAC52902.1; JOINED.
DR   EMBL; U03715; AAC52903.1; -.
DR   EMBL; U03716; AAC52903.1; JOINED.
DR   EMBL; U03718; AAC52903.1; JOINED.
DR   EMBL; U34607; AAC52903.1; JOINED.
DR   EMBL; U34608; AAC52903.1; JOINED.
DR   EMBL; U34609; AAC52903.1; JOINED.
DR   EMBL; U34610; AAC52903.1; JOINED.
DR   EMBL; U34611; AAC52903.1; JOINED.
DR   EMBL; U34612; AAC52903.1; JOINED.
```

```
DR   EMBL; U34613; AAC52903.1; JOINED.
DR   EMBL; U11636; AAC52178.1; -.
DR   EMBL; U11637; AAC52179.1; -.
DR   EMBL; L22545; AAA19787.1; -.
DR   EMBL; AF257775; AAF69009.1; -.
DR   PIR; A56101; A56101.
DR   PDB; 1KOE; 16-FEB-99.
DR   PDB; 1DY0; 11-APR-00.
DR   PDB; 1DY1; 21-JAN-01.
DR   MGD; MGI:88451; Col18a1.
DR   GO; GO:0005604; C:basement membrane; IDA.
DR   GO; GO:0001525; P:angiogenesis; IMP.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 8.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
DR   PROSITE; PS50038; FZ; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;
KW   Alternative promoter usage; 3D-structure.
FT   SIGNAL       1     26       POTENTIAL.
FT   CHAIN       27   1774       COLLAGEN ALPHA 1(XVIII) CHAIN.
FT   CHAIN     1591   1774       ENDOSTATIN.
FT   DOMAIN     245    433       TSP N-TERMINAL.
FT   DOMAIN     365    482       FZ.
FT   DOMAIN      27    785       NONHELICAL REGION 1 (NC1).
FT   DOMAIN     786    812       TRIPLE-HELICAL REGION 1 (COL1).
FT   DOMAIN     813    822       NONHELICAL REGION 2 (NC2).
FT   DOMAIN     823    896       TRIPLE-HELICAL REGION 2 (COL2).
FT   DOMAIN     897    920       NONHELICAL REGION 3 (NC3).
FT   DOMAIN     921   1042       TRIPLE-HELICAL REGION 3 (COL3).
FT   DOMAIN    1043   1065       NONHELICAL REGION 4 (NC4).
FT   DOMAIN    1066   1148       TRIPLE-HELICAL REGION 4 (COL4).
FT   DOMAIN    1149   1162       NONHELICAL REGION 5 (NC5).
FT   DOMAIN    1163   1204       TRIPLE-HELICAL REGION 5 (COL5).
FT   DOMAIN    1205   1217       NONHELICAL REGION 6 (NC6).
FT   DOMAIN    1218   1290       TRIPLE-HELICAL REGION 6 (COL6).
FT   DOMAIN    1291   1300       NONHELICAL REGION 7 (NC7).
FT   DOMAIN    1301   1333       TRIPLE-HELICAL REGION 7 (COL7).
FT   DOMAIN    1334   1345       NONHELICAL REGION 8 (NC8).
FT   DOMAIN    1346   1369       TRIPLE-HELICAL REGION 8 (COL8).
FT   DOMAIN    1370   1376       NONHELICAL REGION 9 (NC9).
FT   DOMAIN    1377   1428       TRIPLE-HELICAL REGION 9 (COL9).
FT   DOMAIN    1429   1441       NONHELICAL REGION 10 (NC10).
FT   DOMAIN    1442   1459       TRIPLE-HELICAL REGION 10 (COL10).
FT   DOMAIN    1460   1774       NONHELICAL REGION 11 (NC11).
FT   CARBOHYD   354    354       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   361    361       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   585    585       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   947    947       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID  1623   1763
FT   DISULFID  1725   1755
FT   SITE      1351   1353       CELL ATTACHMENT SITE (POTENTIAL).
FT   VARSPLIC     1    459       Missing (in isoform 3).
FT                               /FTId=VSP_001157.
FT   VARSPLIC   460    486       AGDRLPVVCASLPSQEDGYCVFIGPAA -> MAPRWHLLDV
FT                               LTSLVLLLVARVSWAEP (in isoform 3).
FT                               /FTId=VSP_001158.
FT   VARSPLIC   240    486       Missing (in isoform 2).
FT                               /FTId=VSP_008303.
FT   CONFLICT  1147   1147       P -> L (IN REF. 4).
FT   CONFLICT  1194   1194       P -> F (IN REF. 4).
FT   CONFLICT  1211   1211       A -> R (IN REF. 4).

  Query Match           87.1%;  Score 778;  DB 1;  Length 1774;
  Best Local Similarity  85.8%;  Pred. No. 5.4e-69;
  Matches  145;  Conservative   13;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||:|||||||||||||||||||||||:|||||||||||||||||||| :||||
Db       1603 VALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRGSVPIV 1662

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              |||||:| |||::|||||:| |:||||||||||:|||||| ||||||||||||:|||| |
Db       1663 NLKDEVLSPSWDSLFSGSQGQLQPGARIFSFDGRDVLRHPAWPQKSVWHGSDPSGRRLME 1722

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 169
              |||||||||   |||||||||| |||| | |||||::||||||||||||:
Db       1723 SYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMTS 1771


RESULT 3
CA1E_HUMAN
ID   CA1E_HUMAN     STANDARD;      PRT;  1388 AA.
AC   P39059;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Collagen alpha 1(XV) chain precursor.
GN   COL15A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Umbilical cord;
RX   MEDLINE=94148920; PubMed=8106446;
RA   Kivirikko S., Heinamaki P., Rehn M.V., Honkanen N., Myers J.C.,
RA   Pihlajaniemi T.;
RT   "Primary structure of the alpha 1 chain of human type XV collagen and
RT   exon-intron organization in the 3' region of the corresponding
RT   gene.";
RL   J. Biol. Chem. 269:4773-4779(1994).
RN   [2]
RP   SEQUENCE OF 1-569 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=94140817; PubMed=8307960;
RA   Muragaki Y., Abe N., Ninomiya Y., Olsen B.R., Ooshima A.;
RT   "The human alpha 1(XV) collagen chain contains a large amino-terminal
RT   non-triple helical domain with a tandem repeat structure and homology
RT   to alpha 1(XVIII) collagen.";
RL   J. Biol. Chem. 269:4042-4046(1994).
RN   [3]
RP   SEQUENCE OF 544-1252 FROM N.A.
RX   MEDLINE=93066196; PubMed=1279671;
RA   Myers J.C., Kivirikko S., Gordon M.K., Dion A.S., Pihlajaniemi T.;
RT   "Identification of a previously unknown human collagen chain, alpha
RT   1(XV), characterized by extensive interruptions in the triple-helical
RT   region.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:10144-10148(1992).
CC   -!- TISSUE SPECIFICITY: Expressed predominantly in internal organs
CC       such as adrenal gland, pancreas and kidney.
CC   -!- PTM: Prolines at the third position of the tripeptide repeating
CC       unit (G-X-Y) are hydroxylated in some or all of the chains.
CC   -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC       INTERRUPTED HELICES (FACIT) FAMILY.
CC   -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L25286; AAA58429.1; -.
DR   EMBL; D21230; BAA04762.1; -.
DR   EMBL; L01697; -; NOT_ANNOTATED_CDS.
```

```
DR   PIR; A53317; A53317.
DR   HSSP; P39061; 1KOE.
DR   Genew; HGNC:2192; COL15A1.
DR   MIM; 120325; -.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 5.
DR   Pfam; PF02210; TSPN; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Cell adhesion; Collagen; Glycoprotein; Signal.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26   1388       COLLAGEN ALPHA 1(XV) CHAIN.
FT   DOMAIN       40    228       TSP N-TERMINAL.
FT   DOMAIN      229    555       NONHELICAL REGION 1 (NC1).
FT   DOMAIN      556    573       TRIPLE-HELICAL REGION 1 (COL1).
FT   DOMAIN      574    618       NONHELICAL REGION 2 (NC2).
FT   DOMAIN      619    732       TRIPLE-HELICAL REGION 2 (COL2).
FT   DOMAIN      733    763       NONHELICAL REGION 3 (NC3).
FT   DOMAIN      764    798       TRIPLE-HELICAL REGION 3 (COL3).
FT   DOMAIN      799    822       NONHELICAL REGION 4 (NC4).
FT   DOMAIN      823    867       TRIPLE-HELICAL REGION 4 (COL4).
FT   DOMAIN      868    878       NONHELICAL REGION 5 (NC5).
FT   DOMAIN      879    949       TRIPLE-HELICAL REGION 5 (COL5).
FT   DOMAIN      950    983       NONHELICAL REGION 6 (NC6).
FT   DOMAIN      984   1013       TRIPLE-HELICAL REGION 6 (COL6).
FT   DOMAIN     1014   1027       NONHELICAL REGION 7 (NC7).
FT   DOMAIN     1028   1045       TRIPLE-HELICAL REGION 7 (COL7).
FT   DOMAIN     1046   1052       NONHELICAL REGION 8 (NC8).
FT   DOMAIN     1053   1107       TRIPLE-HELICAL REGION 8 (COL8).
FT   DOMAIN     1108   1117       NONHELICAL REGION 9 (NC9).
FT   DOMAIN     1118   1132       TRIPLE-HELICAL REGION 9 (COL9).
FT   DOMAIN     1133   1388       NONHELICAL REGION 10 (NC10).
FT   DOMAIN      358    555       4 X TANDEM REPEATS.
FT   REPEAT      358    408       1.
FT   REPEAT      409    459       2.
FT   REPEAT      460    509       3.
FT   REPEAT      510    555       4.
FT   CARBOHYD    306    306       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    324    324       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    687    687       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    807    807       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    814    814       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1046   1046       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     10     10       C -> S (IN REF. 2).
FT   CONFLICT     49     49       D -> V (IN REF. 2).
FT   CONFLICT     95     95       L -> A (IN REF. 2).
FT   CONFLICT    150    150       P -> A (IN REF. 2).
FT   CONFLICT    204    204       M -> V (IN REF. 2).
FT   CONFLICT    409    409       R -> A (IN REF. 2).
SQ   SEQUENCE   1388 AA;  141930 MW;  60822AD925A3093D CRC64;

  Query Match             56.4%;  Score 504;  DB 1;  Length 1388;
  Best Local Similarity   56.9%;  Pred. No. 6.6e-42;
  Matches   95;  Conservative   27;  Mismatches   41;  Indels    4;  Gaps    1;

Qy          2 ALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIVN 61
              ||| | || :|    ||||||:|||| ||  |:|||||| |||| :|||:|:| ::|||| 
Db       1222 ALNMPFSGDIR----ADFQCFKQARAAGLLSTYRAFLSSHLQDLSTIVRKAERYSLPIVN 1277

Qy         62 LKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTES 121
              || ::|| :|:::||| |       |:||||:|:: |:|||| :|||| |:| || :: 
Db       1278 LKGQVLFNNWDSIFSGHGGQFNMHIPIYSFDGRDIMTDPSWPQKVIWHGSSPHGVRLVDN 1337

Qy        122 YCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 168
              ||| |||   : || || |  |::| | | || :  ||||||||||
Db       1338 YCEAWRTADTAVTGLASPLSTGKILDQKAYSCANRLIVLCIENSFMT 1384


RESULT 4
GLI_CHICK
ID   GLI_CHICK      STANDARD;      PRT;   556 AA.
AC   P55878;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein GLI1 (GLI) (Fragment).
GN   GLI1 OR GLI.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97105842; PubMed=8948590;
RA   Marigo V., Johnson R.L., Vortkamp A., Tabin C.J.;
RT   "Sonic hedgehog differentially regulates expression of GLI and GLI3
RT   during limb development.";
RL   Dev. Biol. 180:273-283(1996).
CC   -!- FUNCTION: MAY REGULATE THE TRANSCRIPTION OF SPECIFIC GENES DURING
CC       NORMAL DEVELOPMENT. MAY PLAY A ROLE IN CRANIOFACIAL DEVELOPMENT
CC       AND DIGITAL DEVELOPMENT, AS WELL AS DEVELOPMENT OF THE CENTRAL
CC       NERVOUS SYSTEM AND GASTROINTESTINAL TRACT. IMPLICATED IN THE
CC       TRANSDUCTION OF SHH SIGNAL (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
CC       PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U60762; AAB51659.1; -.
DR   HSSP; P08151; 2GLI.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 5.
DR   SMART; SM00355; ZnF_C2H2; 5.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 5.
KW   Zinc-finger; Metal-binding; DNA-binding; Transcription regulation;
KW   Nuclear protein; Repeat.
FT   ZN_FING     247    272       C2H2-TYPE.
FT   ZN_FING     280    307       C2H2-TYPE.
FT   ZN_FING     313    337       C2H2-TYPE.
FT   ZN_FING     343    368       C2H2-TYPE.
FT   ZN_FING     374    399       C2H2-TYPE.
FT   NON_TER     556    556
SQ   SEQUENCE   556 AA;  60215 MW;  722D2AA5A1CA4D98 CRC64;

  Query Match              8.5%;  Score 76;  DB 1;  Length 556;
  Best Local Similarity   23.2%;  Pred. No. 7.3;
  Matches   48;  Conservative   21;  Mismatches   66;  Indels   72;  Gaps   10;

Qy          6 PLSG---GMRGIRGADFQ-C----------------------------FQQARAVGLAG 32
              || |   |  |::| ||  |                            |  |  | |
Db         20 PLHGASAGTPGLQGLDFPVCHQPNLASSHHGYGLVPGTEHPGGAADGSRFSTPRGAGKLG 79

Qy         33 TFRAF----LSSRLQDLYSIVRRADRAAVPIVN------------LKDELLFPS------ 70
                ||     ||     || :::| :  : |  :|            |    : ||
Db         80 KKRALSISPLSDSSVDLQTVIRTSPNSLVAFINSRCASAGGSYGHLSISTISPSLGYQNP 139

Qy         71 ------WEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRR-----LT 119
                   ||| :  || | :   :   :  | |||  : :: |       |     ||
Db        140 PGQQKGQGQLFSHTP-PLPPCSSHETLSSRPGLLHPTPARGTIKHCQQLKLERSLSSPLT 198

Qy        120 ESYCETWRTE----APSATGQASSLLG 142
              | |  ::|    :|::||    |||
Db        199 AKYPEE-KSEGDISSPASTGTQDPLLG 224
```

```
RESULT 5
XA5L_MOUSE
ID   XA5L_MOUSE     STANDARD;      PRT;   334 AA.
AC   Q9WTJ8; Q80X84; Q9D993;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   XAP-5-like protein.
GN   X5L OR D0H6S2654E.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=20005935; PubMed=10534398;
RA   Sedlack Z., Muenstermann E., Dhorne-Pollet S., Otto C., Bock D.,
RA   Schuetz G., Poustka A.;
RT   "Human and mouse XAP-5 and XAP-5-like (X5L) genes: identification of
RT   an ancient functional retroposon differentially expressed in testis.";
RL   Genomics 61:125-132(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=22354683; PubMed=12466851;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA   Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA   Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA   Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA   Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA   Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA   Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA   Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA   Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA   Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA   Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA   Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA   Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA   Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA   Birney E., Hayashizaki Y.;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- TISSUE SPECIFICITY: Widely expressed. Abundant in testis, where it
CC       is expressed in seminiferous tubules, not in the interstitium. At
CC       the cellular level, expressed in primary spermatocytes and round
CC       spermatids, but not detectable in spermatogonia, elongating
CC       spermatids, mature spermatozoa, Sertoli cells or Leydig cells.
CC   -!- SIMILARITY: Belongs to the XAP5 family.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC       frameshift in position 303.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y18506; CAB46283.1; -.
DR   EMBL; Y18508; CAB46284.1; -.
DR   EMBL; AK007247; BAB24914.1; ALT_FRAME.
DR   EMBL; BC049659; AAH49659.1; -.
DR   MGD; MGI:1351640; D0H6S2654E.
DR   InterPro; IPR007005; XAP5.
DR   Pfam; PF04921; XAP5; 1.
FT   CONFLICT    142    142       A -> V (IN REF. 3).
FT   CONFLICT    229    229       L -> V (IN REF. 2).
SQ   SEQUENCE   334 AA;  39594 MW;  4A438C52BC4A97C6 CRC64;

  Query Match             8.5%;  Score 75.5;  DB 1;  Length 334;
  Best Local Similarity   24.8%;  Pred. No. 4.5;
  Matches   35;  Conservative   13;  Mismatches   38;  Indels   55;  Gaps    7;

Qy          8 SGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAV-PIVNLKDEL 66
              ||  | :|             :    | : ||   || |     |    | |  :: :|::|
Db        204 SGHRRTVR-----------MSKGSTVQQFLKRALQGLRRDFRELRAAGVEQLMYVKEDL 251

Qy         67 LFPSWEALFS-------GSEGPLKPGARIFSFD-------------------GKDVLRH 99
              : | :   :         |  |||      ||||                  || |||
Db        252 ILPHYHTFYDFIVAKARGKSGPL------FSFDVHDDVRLLSDATMEKDESHAGKVVLR- 304

Qy        100 PTWPQK-------SVWHGSDP 113
               :| :|       | |   ||
Db        305 -SWYEKNKHIFPASRWEPYDP 324


RESULT 6
Y4BG_RHISN
ID   Y4BG_RHISN     STANDARD;      PRT;   271 AA.
AC   P55374;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Hypothetical protein Y4BG precursor.
GN   Y4BG.
OS   Rhizobium sp. (strain NGR234).
OG   Plasmid sym pNGR234a.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX   NCBI_TaxID=394;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97305956; PubMed=9163424;
RA   Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA   Perret X.;
RT   "Molecular basis of symbiosis between Rhizobium and legumes.";
```

```
RL   Nature 387:394-401(1997).
CC   -!- SIMILARITY: None obvious.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000066; AAB91622.1; -.
KW   Hypothetical protein; Plasmid; Signal.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    271       HYPOTHETICAL PROTEIN Y4BG.
SQ   SEQUENCE   271 AA;  30034 MW;  5572CF1461FCDA63 CRC64;

  Query Match             8.4%;  Score 75;  DB 1;  Length 271;
  Best Local Similarity   21.1%;  Pred. No. 4;
  Matches   28;  Conservative   26;  Mismatches   35;  Indels   44;  Gaps    7;

Qy          4 NSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRA------DRAAV 57
              |: ::  :: : ||||: ||       :| |::       | | | |        :: |:
Db        174 NADIAATIKSLLGADFEAFQAIITGTGSGEFKS-------DDY-IGRTCTPHMCREQEAL 225

Qy         58 PIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHP---TWPQKSVWHGSDPN 114
                :: ||   : :|          ||         | :: ||    ||:|:
Db        226 LFLSAKDRRAYAAW----------KP-------HQKKIIVHPPVKQWPEKA-------- 259

Qy        115 GRRLTESYCETWR 127
                ::   :: |||:
Db        260 -KQELRAWAETWK 271


RESULT 7
CRK_XENLA
ID   CRK_XENLA       STANDARD;       PRT;   296 AA.
AC   P87378;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   SH2/SH3 adaptor crk (Adapter molecule crk) (CRK2).
GN   CRK.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97180926; PubMed=9029144;
RA   Evans E.K., Lu W., Strum S.L., Mayer B.J., Kornbluth S.;
RT   "Crk is required for apoptosis in Xenopus egg extracts.";
RL   EMBO J. 16:230-241(1997).
CC   -!- FUNCTION: Required for apoptosis in Xenopus egg extracts. May
CC       mediate attachment-induced MAPK8 activation, membrane ruffling and
CC       cell motility in a Rac-dependent manner.
CC   -!- SUBUNIT: Interacts with ABL, C3G, SOS, MAP4K1, MAPK8 and DOCK1 via
CC       its first SH3 domain. Interacts with BCAR1, CBL, PXN and GAB1 via
CC       its SH2 domain upon stimulus-induced tyrosine phosphorylation.
CC       Interacts with several tyrosine-phosphorylated growth factor
CC       receptors such as EGFR, PDGFR and INSR via its SH2 domain (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic; translocated to the plasma
CC       membrane upon cell adhesion (By similarity).
CC   -!- PTM: Phosphorylated on Tyr-214 upon cell adhesion. Results in the
CC       negative regulation of the association with SH2- and SH3-binding
CC       partners, possibly by the formation of an intramolecular
CC       interaction of phosphorylated Tyr-221 with the SH2 domain. This
CC       leads finally to the down-regulation of the Crk signaling pathway
CC       (By similarity).
CC   -!- SIMILARITY: Contains 1 SH2 domain.
CC   -!- SIMILARITY: Contains 2 SH3 domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U89774; AAB49698.1; -.
DR   HSSP; Q64010; 1CKA.
DR   InterPro; IPR000980; SH2.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00017; SH2; 1.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00401; SH2DOMAIN.
DR   PRINTS; PR00452; SH3DOMAIN.
DR   ProDom; PD000093; SH2; 1.
DR   ProDom; PD000066; SH3; 1.
DR   SMART; SM00252; SH2; 1.
DR   SMART; SM00326; SH3; 2.
DR   PROSITE; PS50001; SH2; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   SH2 domain; SH3 domain; Repeat; Phosphorylation.
FT   DOMAIN       13    112       SH2.
FT   DOMAIN      125    185       SH3 1.
FT   DOMAIN      249    289       SH3 2.
FT   MOD_RES     214    214       PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   296 AA;  33409 MW;  544F11F4A1F75A66 CRC64;

  Query Match             8.3%;  Score 74;  DB 1;  Length 296;
  Best Local Similarity   32.4%;  Pred. No. 5.5;
  Matches   24;  Conservative   12;  Mismatches   20;  Indels   18;  Gaps    5;

Qy         75 FSGSEG---PLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRR--LTESYCETWRTE 129
                | |::    | | |          |:||     |::  |:  | :|||  :   | | :|
Db        136 FIGNDDEDLPFKKG---------DILRIREKPEEQWWNAEDSDGRRGMIPVPYVEKYR-- 184

Qy        130 APSATGQASSLLGG 143
                ||: |  |:|:||
Db        185 PPSSPG--SALIGG 196


RESULT 8
STRN_RAT
ID   STRN_RAT       STANDARD;       PRT;   780 AA.
AC   P70483;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Striatin.
GN   STRN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 222-243; 258-269; 451-456; 488-494
RP   AND 764-777.
RC   STRAIN=Wistar; TISSUE=Brain;
RX   MEDLINE=96354919; PubMed=8769426;
RA   Castets F., Bartoli M., Barnier J.V., Baillat G., Salin P.,
RA   Moqrich A., Bourgeois J.-P., Denizot F., Rougon G., Calothy G.,
RA   Monneron A.;
RT   "A novel calmodulin-binding protein, belonging to the WD-repeat
RT   family, is localized in dendrites of a subset of CNS neurons.";
RL   J. Cell Biol. 134:1051-1062(1996).
CC   -!- FUNCTION: Binds calmodulin in a calcium dependent manner. May
CC       function as scaffolding or signaling protein.
CC   -!- SUBUNIT: INTERACTS WITH PROTEIN PHOSPHATASE 2A (PP2A) (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic and membrane-bound.
CC   -!- TISSUE SPECIFICITY: MAINLY EXPRESSED IN THE CENTRAL NERVOUS
CC       SYSTEM. MOSTLY CONFINED IN DENDRITES, NOT IN AXONS, AND IS MOST
```

```
CC       ABUNDANT IN DENTRITIC SPINES.
CC   -!- SIMILARITY: Belongs to the WD-repeat striatin family.
CC   -!- SIMILARITY: Contains 6 WD repeats.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X99326; CAA67701.1; -.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF00400; WD40; 6.
DR   PRINTS; PR00320; GPROTEINBRPT.
DR   ProDom; PD000018; WD40; 2.
DR   SMART; SM00320; WD40; 6.
DR   PROSITE; PS00678; WD_REPEATS_1; 2.
DR   PROSITE; PS50082; WD_REPEATS_2; 4.
DR   PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW   Calmodulin-binding; Repeat; WD repeat; Coiled coil.
FT   DOMAIN       53    120       COILED COIL (POTENTIAL).
FT   DOMAIN      149    166       CALMODULIN-BINDING (POTENTIAL).
FT   REPEAT      461    500       WD 1.
FT   REPEAT      514    553       WD 2.
FT   REPEAT      567    606       WD 3.
FT   REPEAT      662    701       WD 4.
FT   REPEAT      704    743       WD 5.
FT   REPEAT      750    779       WD 6.
FT   SITE         55     63       CAVEOLIN-BINDING (POTENTIAL).
SQ   SEQUENCE   780 AA;  86226 MW;  67F7C22099D560F8 CRC64;

  Query Match             8.3%;  Score 74;  DB 1;  Length 780;
  Best Local Similarity   23.7%;  Pred. No. 17;
  Matches   51;  Conservative   19;  Mismatches   65;  Indels   80;  Gaps   12;

Qy          2 ALNSPLSGGMRGIRGADFQCFQQARAVG-LAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||  | | |   | ||| :  :    :| |||              ::   ||  |  |
Db        398 ALTFPPSSGKSFIMGAD-EALESELGLGELAG-------------LTVANEADSLAYDIA 443

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLR-HPTWP---------------- 103
              | || |   :|   |:     |:       |||   |  ||  |
Db        444 NNKDALR-KTWNPKFT-----LRS-----HFDGIRALAFHPIEPVLITASEDHTLKMWNL 492

Qy        104 ------QKSVWHGSDP-----------------NGRRL----TESYCETWRTEAPSATG 135
                   :||    :|                  || :     |:   ::| |  |:
Db        493 QKTAPAKKSTSLDVEPIYTFRAHKGPVLCVVMSSNGEQCYSGGTDGLIQSWSTTNPNVDP 552

Qy        136 QAS---SLLGGRLLGQS------AASCHHAYIVLC 161
                |   |:| | ||| :      | |  |  :: |
Db        553 YDSYDPSVLRGPLLGHTDAVWGLAYSAAHQRLLSC 587


RESULT 9
XA5L_HUMAN
ID   XA5L_HUMAN     STANDARD;      PRT;   325 AA.
AC   Q9Y247;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   XAP-5-like protein.
GN   X5L.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20005935; PubMed=10534398;
RA   Sedlack Z., Muenstermann E., Dhorne-Pollet S., Otto C., Bock D.,
RA   Schuetz G., Poustka A.;
RT   "Human and mouse XAP-5 and XAP-5-like (X5L) genes: identification of
RT   an ancient functional retroposon differentially expressed in testis.";
RL   Genomics 61:125-132(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cervix;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- TISSUE SPECIFICITY: Widely expressed. Mostly abundant in testis
CC       and adult and fetal brain.
CC   -!- SIMILARITY: Belongs to the XAP5 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y18504; CAB46273.1; -.
DR   EMBL; Y18503; CAB46272.1; -.
DR   EMBL; BC001261; AAH01261.1; -.
DR   InterPro; IPR007005; XAP5.
DR   Pfam; PF04921; XAP5; 1.
SQ   SEQUENCE   325 AA;  38708 MW;  9F0653C3D7236997 CRC64;

  Query Match             8.2%;  Score 73.5;  DB 1;  Length 325;
  Best Local Similarity   24.8%;  Pred. No. 6.9;
  Matches   35;  Conservative   12;  Mismatches   39;  Indels   55;  Gaps    7;

Qy          8 SGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAV-PIVNLKDEL 66
              ||  | :|              |    | : ||   || |          | |  :: :|::|
Db        195 SGHRRTVR------------VRKGNTVQQFLKKALQGLRKDFLELRSAGVEQLMFIKEDL 242

Qy         67 LFPSWEALFS-------GSEGPLKPGARIFSFD-------------------GKDVLRH 99
              : | :   :         |  |||       ||||                  || |||
Db        243 ILPHYHTFYDFIIARARGKSGPL------FSFDVHDDVRLLSDATMEKDESHAGKVVLR- 295

Qy        100 PTWPQK-------SVWHGSDP 113
               :| :|       | |   ||
Db        296 -SWYEKNKHIFPASRWEAYDP 315


RESULT 10
YFHM_ECOLI
ID   YFHM_ECOLI     STANDARD;      PRT;  1653 AA.
AC   P76578;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical lipoprotein yfhM precursor.
GN   YFHM OR B2520.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
```

```
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC       (Potential).
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE000338; AAC75573.1; -.
DR   PIR; G65028; G65028.
DR   EcoGene; EG13394; yfhM.
DR   InterPro; IPR000437; Prok_lipoprot_S.
DR   InterPro; IPR008930; Terp_cyc_toroid.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Hypothetical protein; Membrane; Lipoprotein; Signal; Coiled coil;
KW   Complete proteome; Palmitate.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   1653       HYPOTHETICAL LIPOPROTEIN YFHM.
FT   DOMAIN     1559   1589       COILED COIL (POTENTIAL).
FT   LIPID        18     18       N-palmitoyl cysteine (Potential).
FT   LIPID        18     18       S-diacylglycerol cysteine (Potential).
SQ   SEQUENCE   1653 AA;  181584 MW;  13109EC5CDEB41A0 CRC64;

  Query Match             8.2%;  Score 73.5;  DB 1;  Length 1653;
  Best Local Similarity   24.4%;  Pred. No. 46;
  Matches   32;  Conservative   21;  Mismatches   47;  Indels   31;  Gaps    5;

Qy         29 GLAGTFRAFLSSRLQDLY---------SIVRRADRAAVPIVNLKDELLFPSWEALFSGSE 79
              |  |  :  |    :|||           ::| ||  |:|   :| ::: |  : | |
Db        380 GAPGYSKQFFMFGPRDLYRPGETVILNGLLRDADGKALPNQPIKLDVIKPDGQVLRSVVS 439

Qy         80 GPLKPGARIFSFDGKDVLRHPTWPQKS-----VWH---GSDPNGRRLTESYCETWRTE-- 129
              |            :: | | |||  |     :||    :  |  |: : : | :  |
Db        440 QP-----------ENGLYHFTWPLDSNAATGMWHIRANTGDNQYRMWDFHVEDFMPERM 487

Qy        130 APSATGQASSL 140
              | : ||: : |
Db        488 ALNLTGEKTPL 498


RESULT 11
PPX_ECOLI
ID   PPX_ECOLI      STANDARD;      PRT;   512 AA.
AC   P29014; P76981;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Exopolyphosphatase (EC 3.6.1.11) (ExopolyPase) (Metaphosphatase).
GN   PPX OR B2502 OR C3020 OR Z3765 OR ECS3364.
OS   Escherichia coli,
OS   Escherichia coli O6, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562, 217992, 83334;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RC   STRAIN=K12;
RX   MEDLINE=93107072; PubMed=8380170;
RA   Akiyama M., Crooke E., Kornberg A.;
RT   "An exopolyphosphatase of Escherichia coli. The enzyme and its ppx
RT   gene in a polyphosphate operon.";
RL   J. Biol. Chem. 268:633-639(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97349980; PubMed=9205837;
RA   Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA   Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA   Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA   Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasundaram S.,
RA   Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA   Yamagata S., Horiuchi T.;
RT   "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT   - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT   analysis of its sequence features.";
RL   DNA Res. 4:91-113(1997).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX   MEDLINE=22388234; PubMed=12471157;
RA   Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA   Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA   Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA   Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT   "Extensive mosaic structure revealed by the complete genome sequence
RT   of uropathogenic Escherichia coli.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
RN   [7]
RP   SIMILARITY TO GPPA.
RX   MEDLINE=94025037; PubMed=8212131;
RA   Reizer J., Reizer A., Saier M.H. Jr., Bork B., Sander C.;
RT   "Exopolyphosphate phosphatase and guanosine pentaphosphate
RT   phosphatase belong to the sugar kinase/actin/hsp 70 superfamily.";
RL   Trends Biochem. Sci. 18:247-248(1993).
CC   -!- FUNCTION: DEGRADATION OF INORGANIC POLYPHOSPHATES. ORTHOPHOSPHATE
CC       IS RELEASED PROGRESSIVELY FROM THE ENDS OF POLYPHOSPHATE OF CIRCA
CC       500 RESIDUES LONG, BUT CHAINS OF CIRCA 15 RESIDUES COMPETE POORLY
```

```
CC       WITH POLYPHOSPHATE AS SUBSTRATE.
CC   -!- CATALYTIC ACTIVITY: {Polyphosphate}(N) + H(2)O =
CC       {polyphosphate}(N-1) + phosphate.
CC   -!- COFACTOR: Magnesium.
CC   -!- SUBUNIT: Homodimer.
CC   -!- SUBCELLULAR LOCATION: Membrane-associated.
CC   -!- SIMILARITY: Belongs to the gppA / ppx family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L06129; AAA24415.1; -.
DR   EMBL; AE000336; AAC75555.1; -.
DR   EMBL; D90878; BAA16390.1; -.
DR   EMBL; D90880; BAA16392.1; -.
DR   EMBL; AE016764; AAN81470.1; -.
DR   EMBL; AE005479; AAG57612.1; -.
DR   EMBL; AP002561; BAB36787.1; -.
DR   PIR; A45333; A45333.
DR   PIR; D91049; D91049.
DR   PIR; H85893; H85893.
DR   EcoGene; EG11403; ppx.
DR   InterPro; IPR003695; Ppx_GppA.
DR   Pfam; PF02541; Ppx-GppA; 1.
KW   Hydrolase; Magnesium; Membrane; Complete proteome.
FT   INIT_MET      0      0
SQ   SEQUENCE   512 AA;  58004 MW;  48611AFF5D9FB9C3 CRC64;

  Query Match             8.2%;  Score 73;  DB 1;  Length 512;
  Best Local Similarity   21.9%;  Pred. No. 13;
  Matches   46;  Conservative   28;  Mismatches   66;  Indels   70;  Gaps   11;

Qy         22 FQQAR--------------------AVGLAGTFRA--------------FLSSRLQDLY 46
              ||:||                    |:| :|| :|                  ||: |
Db        186 FQRARMAAAQKLETLTWQFRIQGWNVAMGASGTIKAAHEVLMEMGEKDGIITPERLEKLV 245

Qy         47 -SIVRRADRAAVPIVNLKDE--LLFPSWEALFSG------------SEGPLKPGARIFSF 91
               ::|  : |::  :  | :|   :|    |:  |            |:| |: |  ::
Db        246 KEVLRHRNFASLSLPGLSEERKTVFVPGLAILCGVFDALAIRELRLSDGALREGV-LYEM 304

Qy         92 DGK----DVLRHPTWPQKSVWHGSDPNGRRL---TESYCETWRTEAPS-ATGQASSLLG- 142
              :|:    ||        : :|      ||:   |    | || : |  |  |  :||
Db        305 EGRFRHQDVRSRTASSLANQYHIDSEQARRVLDTTMQMYEQWREQQPKLAHPQLEALLRW 364

Qy        143 -------GRLLGQSAASCHHAYIVLCIENS 165
                    |  :  |    | |||   ::||
Db        365 AAMLHEVGLNINHSGLHRHSAYI---LQNS 391


RESULT 12
YP67_MYCTU
ID   YP67_MYCTU     STANDARD;      PRT;   884 AA.
AC   Q50654; Q50731;
DT   01-NOV-1997 (Rel. 35, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical protein Rv2567/MT2643.
GN   RV2567 OR MT2643 OR MTCY227.34C OR MTCY9C4.01C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RX   MEDLINE=22206494; PubMed=12218036;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT   "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   J. Bacteriol. 184:5479-5490(2002).
CC   -!- SIMILARITY: SOME, TO M.TUBERCULOSIS RV2411C AND SYNECHOCYSTIS PCC
CC       6803 SLL0335.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z77250; CAB01053.1; -.
DR   EMBL; AE007098; AAK46956.1; -.
DR   PIR; C70729; C70729.
DR   TIGR; MT2643; -.
DR   TubercuList; Rv2567; -.
DR   InterPro; IPR007296; DUF403.
DR   InterPro; IPR007297; DUF404.
DR   InterPro; IPR007302; DUF407.
DR   Pfam; PF04168; DUF403; 1.
DR   Pfam; PF04169; DUF404; 1.
DR   Pfam; PF04174; DUF407; 1.
KW   Hypothetical protein; Complete proteome.
FT   CONFLICT    645    645       Q -> R (IN REF. 2).
SQ   SEQUENCE   884 AA;  95448 MW;  95D23A4D2EDEB365 CRC64;

  Query Match             8.1%;  Score 72.5;  DB 1;  Length 884;
  Best Local Similarity   29.5%;  Pred. No. 28;
  Matches   31;  Conservative   15;  Mismatches   26;  Indels   33;  Gaps    6;

Qy         51 RADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDG--------KDVLRHPTW 102
              |||  ||            ||   |:| :  | :||: : | :|        :| | : ||
Db        609 RADMIAVA----------PS--TLWSLTVDPDRPGSLVQSVEGLALAAQAVRDQLSNDTW 656

Qy        103 -----PQKSVWHGSDPNGRRLTESYCETWRTEAPSATGQASSLLG 142
                  :::| | |||      :|  |    :|  |: || :| |
Db        657 MVLANVERAVEHKSDP-----PQSLAE---ADAVLASAQAETLAG 693


RESULT 13
YP97_MYCBO
ID   YP97_MYCBO     STANDARD;      PRT;   884 AA.
AC   P59974;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Hypothetical protein Mb2597.
GN   MB2597.
OS   Mycobacterium bovis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1765;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AF2122/97;
RX   MEDLINE=22709107; PubMed=12788972;
RA   Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA   Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
RA   Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA   Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT   "The complete genome sequence of Mycobacterium bovis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; BX248343; CAD94782.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   884 AA;  95476 MW;  FCF6A80543686808 CRC64;

  Query Match             8.1%;  Score 72.5;  DB 1;  Length 884;
  Best Local Similarity   29.5%;  Pred. No. 28;
  Matches   31;  Conservative   15;  Mismatches   26;  Indels   33;  Gaps    6;

Qy         51 RADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDG--------KDVLRHPTW 102
              |||  ||             ||   |:|  :  | :||: : | :|         :| | : ||
Db        609 RADMIAVA----------PS--TLWSLTVDPDRPGSLVQSVEGLALAARAVRDQLSNDTW 656

Qy        103 -----PQKSVWHGSDPNGRRLTESYCETWRTEAPSATGQASSLLG 142
                  :::| | |||      :|  |    :|  |: || :| |
Db        657 MVLANVERAVEHKSDP-----PQSLAE---ADAVLASAQAETLAG 693


RESULT 14
CRK_HUMAN
ID   CRK_HUMAN      STANDARD;      PRT;   304 AA.
AC   P46108;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Proto-oncogene C-crk (P38) (Adapter molecule crk).
GN   CRK.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryonic lung, and Placenta;
RX   MEDLINE=92334347; PubMed=1630456;
RA   Matsuda M., Tanaka S., Nagata S., Kojima A., Kurata T., Shibuya M.;
RT   "Two species of human CRK cDNA encode proteins with distinct
RT   biological activities.";
RL   Mol. Cell. Biol. 12:3482-3489(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93390962; PubMed=8378094;
RA   Fioretos T., Heisterkamp N., Groffen J., Benjes S., Morris C.;
RT   "CRK proto-oncogene maps to human chromosome band 17p13.";
RL   Oncogene 8:2853-2855(1993).
RN   [3]
RP   INTERACTION WITH DOCK1.
RX   MEDLINE=96239533; PubMed=8657152;
RA   Hasegawa H., Kiyokawa E., Tanaka S., Nagashima K., Gotoh N.,
RA   Shibuya M., Kurata T., Matsuda M.;
RT   "DOCK180, a major CRK-binding protein, alters cell morphology upon
RT   translocation to the cell membrane.";
RL   Mol. Cell. Biol. 16:1770-1776(1996).
RN   [4]
RP   INTERACTION WITH DOCK1; C3G AND EPS15, AND MUTAGENESIS OF
RP   ASP-150.
RX   MEDLINE=96278913; PubMed=8662907;
RA   Matsuda M., Ota S., Tanimura R., Nakamura H., Matuoka K., Takenawa T.,
RA   Nagashima K., Kurata T.;
RT   "Interaction between the amino-terminal SH3 domain of CRK and its
RT   natural target proteins.";
RL   J. Biol. Chem. 271:14468-14472(1996).
RN   [5]
RP   INTERACTION WITH DOCK4.
RX   MEDLINE=22515525; PubMed=12628187;
RA   Yajnik V., Paulding C., Sordella R., McClatchey A.I., Saito M.,
RA   Wahrer D.C.R., Reynolds P., Bell D.W., Lake R., van den Heuvel S.,
RA   Settleman J., Haber D.A.;
RT   "DOCK4, a GTPase activator, is disrupted during tumorigenesis.";
RL   Cell 112:673-684(2003).
CC   -!- FUNCTION: The Crk-I and Crk-II forms differ in their biological
CC       activities. Crk-II has less transforming activity than Crk-I. Crk-
CC       II mediates attachment-induced MAPK8 activation, membrane ruffling
CC       and cell motility in a Rac-dependent manner. Involved in
CC       phagocytosis of apoptotic cells and cell motility via its
CC       interaction with DOCK1 and DOCK4.
CC   -!- SUBUNIT: Interacts with ABL, C3G, SOS, MAP4K1, MAPK8 and DOCK3 via
CC       its first SH3 domain. Interacts with BCAR1, CBL, PXN and GAB1 via
CC       its SH2 domain upon stimulus-induced tyrosine phosphorylation.
CC       Interacts with several tyrosine-phosphorylated growth factor
CC       receptors such as EGFR, PDGFR and INSR via its SH2 domain (By
CC       similarity). Interacts with DOCK1 and DOCK4.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic; translocated to the plasma
CC       membrane upon cell adhesion (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Crk-II;
CC         IsoId=P46108-1; Sequence=Displayed;
CC       Name=Crk-I;
CC         IsoId=P46108-2; Sequence=VSP_004173;
CC   -!- DOMAIN: The C-terminal SH3 domain function as a negative modulator
CC       for transformation and the N-terminal SH3 domain appears to
CC       function as a positive regulator for transformation (By
CC       similarity).
CC   -!- PTM: Phosphorylation of Crk-II (40 kDa) gives rise to a 42 kda
CC       form.
CC   -!- PTM: Phosphorylated on Tyr-221 upon cell adhesion. Results in the
CC       negative regulation of the association with SH2- and SH3-binding
CC       partners, possibly by the formation of an intramolecular
CC       interaction of phosphorylated Tyr-221 with the SH2 domain. This
CC       leads finally to the down-regulation of the Crk signaling pathway
CC       (By similarity).
CC   -!- SIMILARITY: Contains 1 SH2 domain.
CC   -!- SIMILARITY: Contains 2 SH3 domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10656; BAA01505.1; -.
DR   EMBL; S65701; AAB28213.1; -.
DR   PIR; A45022; A45022.
DR   PDB; 1JU5; 06-NOV-02.
DR   SWISS-2DPAGE; P46108; HUMAN.
DR   Genew; HGNC:2362; CRK.
DR   MIM; 164762; -.
DR   GO; GO:0005737; C:cytoplasm; TAS.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   GO; GO:0030036; P:actin cytoskeleton organization and biogenesis; TAS.
DR   GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.
DR   GO; GO:0007165; P:signal transduction; TAS.
DR   InterPro; IPR000980; SH2.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00017; SH2; 1.
```

```
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00401; SH2DOMAIN.
DR   PRINTS; PR00452; SH3DOMAIN.
DR   ProDom; PD000093; SH2; 1.
DR   ProDom; PD000066; SH3; 1.
DR   SMART; SM00252; SH2; 1.
DR   SMART; SM00326; SH3; 2.
DR   PROSITE; PS50001; SH2; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Proto-oncogene; SH2 domain; SH3 domain; Repeat; Alternative splicing;
KW   Phosphorylation; 3D-structure.
FT   DOMAIN       13    118       SH2.
FT   DOMAIN      132    192       SH3 1.
FT   DOMAIN      256    296       SH3 2.
FT   MOD_RES     221    221       PHOSPHORYLATION (BY SIMILARITY).
FT   VARSPLIC    205    304       Missing (in isoform Crk-I).
FT                                /FTId=VSP_004173.
FT   MUTAGEN     150    150       D->K: ABOLISHES INTERACTION WITH DOCK1.
SQ   SEQUENCE   304 AA;  33872 MW;  D74A83ED1FFC0EBC CRC64;

  Query Match             8.1%;  Score 72;  DB 1;  Length 304;
  Best Local Similarity   27.8%;  Pred. No. 9;
  Matches   20;  Conservative   12;  Mismatches   24;  Indels   16;  Gaps    3;

Qy         88 IFSFDGK----------DVLRHPTWPQKSVWHGSDPNGRR--LTESYCETWRTEAPSATG 135
              :| |:|           |:||    |:: |:  |  |:|  :   | | :|     |:
Db        140 LFDFNGNDEEDLPFKKGDILRIRDKPEEQWWNAEDSEGKRGMIPVPYVEKYR----PASA 195

Qy        136 QASSLLGGRLLG 147
                |:|:||   |
Db        196 SVSALIGGNQEG 207


RESULT 15
CRK_MOUSE
ID   CRK_MOUSE      STANDARD;      PRT;   304 AA.
AC   Q64010;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Proto-oncogene C-crk (P38) (Adapter molecule crk).
GN   CRK OR CRKO.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS CRK-I AND CRK-II).
RC   TISSUE=Liver;
RX   MEDLINE=94239744; PubMed=8183562;
RA   Ogawa S., Toyoshima H., Kozutsumi H., Hagiwara K., Sakai R.,
RA   Tanaka T., Hirano N., Mano H., Yazaki Y., Hirai H.;
RT   "The C-terminal SH3 domain of the mouse c-Crk protein negatively
RT   regulates tyrosine-phosphorylation of Crk associated p130 in rat 3Y1
RT   cells.";
RL   Oncogene 9:1669-1678(1994).
RN   [2]
RP   INTERACTION WITH REPS1.
RC   TISSUE=Muscle;
RX   MEDLINE=98058900; PubMed=9395447;
RA   Yamaguchi A., Urano T., Goi T., Feig L.A.;
RT   "An eps homology (EH) domain protein that binds to the ral-GTPase
RT   target, RalBP1.";
RL   J. Biol. Chem. 272:31230-31234(1997).
RN   [3]
RP   INTERACTION WITH DDEF1/ASAP1.
RX   MEDLINE=99038209; PubMed=9819391;
RA   Brown M.T., Andrade J., Radhakrishna H., Donaldson J.G., Cooper J.A.,
RA   Randazzo P.A.;
RT   "ASAP1, a phospholipid-dependent arf GTPase-activating protein that
RT   associates with and is phosphorylated by Src.";
RL   Mol. Cell. Biol. 18:7038-7051(1998).
RN   [4]
RP   INTERACTION WITH DOCK3.
RC   TISSUE=Brain;
RX   MEDLINE=20312861; PubMed=10854253;
RA   Kashiwa A., Yoshida H., Lee S., Paladino T., Liu Y., Chen Q.,
RA   Dargusch R., Schubert D., Kimura H.;
RT   "Isolation and characterization of novel presenilin binding
RT   protein.";
RL   J. Neurochem. 75:109-116(2000).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 134-190.
RX   MEDLINE=95253821; PubMed=7735837;
RA   Wu X., Knudsen B., Feller S.M., Zheng J., Sali A., Cowburn D.,
RA   Hanafusa H., Kuriyan J.;
RT   "Structural basis for the specific interaction of lysine-containing
RT   proline-rich peptides with the N-terminal SH3 domain of c-Crk.";
RL   Structure 3:215-226(1995).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 133-191.
RX   MEDLINE=99069628; PubMed=9851931;
RA   Nguyen J.T., Turck C.W., Cohen F.E., Zuckermann R.N., Lim W.A.;
RT   "Exploiting the basis of proline recognition by SH3 and WW domains:
RT   design of N-substituted inhibitors.";
RL   Science 282:2088-2092(1998).
CC   -!- FUNCTION: The Crk-I and Crk-II forms differ in their biological
CC       activities. Crk-II has less transforming activity than Crk-I. Crk-
CC       II mediates attachment-induced MAPK8 activation, membrane ruffling
CC       and cell motility in a Rac-dependent manner.
CC   -!- SUBUNIT: Interacts with ABL, C3G, SOS, MAP4K1 and MAPK8 via its
CC       first SH3 domain. Interacts with BCAR1, CBL, PXN and GAB1 via its
CC       SH2 domain upon stimulus-induced tyrosine phosphorylation.
CC       Interacts with several tyrosine-phosphorylated growth factor
CC       receptors such as EGFR, PDGFR and INSR via its SH2 domain.
CC       Interacts with DOCK1, DOCK4, C3G and EPS15 via its SH3 domain (By
CC       similarity). Interacts with DOCK3. Interacts with REPS1 and
CC       DDEF1/ASAP1 via its SH3 domain.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic; translocated to the plasma
CC       membrane upon cell adhesion (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Crk-II;
CC         IsoId=Q64010-1; Sequence=Displayed;
CC       Name=Crk-I;
CC         IsoId=Q64010-2; Sequence=VSP_004174;
CC   -!- TISSUE SPECIFICITY: Ubiquitous.
CC   -!- DOMAIN: The C-terminal SH3 domain function as a negative modulator
CC       for transformation and the N-terminal SH3 domain appears to
CC       function as a positive regulator for transformation (By
CC       similarity).
CC   -!- PTM: Phosphorylated on Tyr-221 upon cell adhesion. Results in the
CC       negative regulation of the association with SH2- and SH3-binding
CC       partners, possibly by the formation of an intramolecular
CC       interaction of phosphorylated Tyr-221 with the SH2 domain. This
CC       leads finally to the down-regulation of the Crk signaling pathway
CC       (By similarity).
CC   -!- SIMILARITY: Contains 1 SH2 domain.
CC   -!- SIMILARITY: Contains 2 SH3 domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S72408; AAB30755.1; -.
DR   PDB; 1CKA; 08-MAY-95.
DR   PDB; 1CKB; 08-MAY-95.
DR   PDB; 1B07; 06-JAN-99.
DR   PDB; 1JU5; 06-NOV-02.
DR   MGD; MGI:88508; Crk.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   InterPro; IPR000980; SH2.
```

```
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00017; SH2; 1.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00401; SH2DOMAIN.
DR   PRINTS; PR00452; SH3DOMAIN.
DR   ProDom; PD000093; SH2; 1.
DR   ProDom; PD000066; SH3; 1.
DR   SMART; SM00252; SH2; 1.
DR   SMART; SM00326; SH3; 2.
DR   PROSITE; PS50001; SH2; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Proto-oncogene; SH2 domain; SH3 domain; Repeat; Alternative splicing;
KW   Phosphorylation; 3D-structure.
FT   DOMAIN       13    118       SH2.
FT   DOMAIN      132    192       SH3 1.
FT   DOMAIN      256    296       SH3 2.
FT   MOD_RES     221    221       PHOSPHORYLATION (BY SIMILARITY).
FT   VARSPLIC    205    304       Missing (in isoform Crk-I).
FT                                /FTId=VSP_004174.
FT   STRAND      136    139
FT   STRAND      143    143
FT   TURN        148    149
FT   STRAND      150    150
FT   STRAND      153    153
FT   TURN        155    156
FT   STRAND      158    163
FT   STRAND      169    173
FT   TURN        175    176
FT   STRAND      179    183
FT   HELIX       184    186
FT   STRAND      187    188
SQ   SEQUENCE   304 AA;  33814 MW;  5491896FC7A89065 CRC64;

  Query Match             8.1%;  Score 72;  DB 1;  Length 304;
  Best Local Similarity   27.8%;  Pred. No. 9;
  Matches   20;  Conservative   12;  Mismatches   24;  Indels   16;  Gaps    3;

Qy         88 IFSFDGK----------DVLRHPTWPQKSVWHGSDPNGRR--LTESYCETWRTEAPSATG 135
              :| |:|           |:||    |::  |:  |  |:|  :   | | :|     |:
Db        140 LFDFNGNDEEDLPFKKGDILRIRDKPEEQWWNAEDSEGKRGMIPVPYVEKYR----PASA 195

Qy        136 QASSLLGGRLLG 147
                |:|:||   |
Db        196 SVSALIGGNQEG 207


Search completed: March 13, 2004, 08:16:59
Job time : 20 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 13, 2004, 08:14:01 ; Search time 42 Seconds
(without alignments)
1277.097 Million cell updates/sec

Title: US-09-171-607A-1
Perfect score: 893
Sequence: 1 VALNSPLSGGMRGIRGADFQ.....ASCHHAYIVLCIENSFMTAS 170

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	893	100.0	187	4	Q8WXI5	Q8wxi5 homo sapien
2	893	100.0	261	4	Q8NG19	Q8ng19 homo sapien
3	893	100.0	816	4	Q8N4S4	Q8n4s4 homo sapien
4	778	87.1	1140	11	Q61434	Q61434 mus musculu
5	763	85.4	226	11	Q9QZD2	Q9qzd2 rattus norv
6	757	84.8	171	11	Q9WUW5	Q9wuw5 rattus norv
7	722	80.9	160	11	Q9CRT2	Q9crt2 mus musculu
8	711	79.6	1344	13	O93419	O93419 gallus gall
9	654	73.2	1315	13	Q8QHL9	Q8qhl9 xenopus lae
10	643	72.0	1307	13	Q8JFF7	Q8jff7 xenopus lae
11	606	67.9	361	13	Q8AWC6	Q8awc6 brachydanio
12	529	59.2	102	4	Q96T70	Q96t70 homo sapien
13	515	57.7	1367	11	O35206	O35206 mus musculu
14	512	57.3	1367	11	Q9EQD9	Q9eqd9 mus musculu
15	505	56.6	1388	4	Q9Y4W4	Q9y4w4 homo sapien
16	427.5	47.9	950	5	Q86SC8	Q86sc8 ciona intes
17	424.5	47.5	299	5	Q8MSE3	Q8mse3 drosophila
18	418.5	46.9	778	5	Q86BH1	Q86bh1 drosophila
19	418.5	46.9	792	5	Q8MT89	Q8mt89 drosophila
20	357	40.0	650	5	O17866	O17866 caenorhabdi
21	357	40.0	778	5	Q9U9K6	Q9u9k6 caenorhabdi
22	357	40.0	1117	5	Q9U9K7	Q9u9k7 caenorhabdi
23	357	40.0	1154	5	Q8I0G6	Q8i0g6 caenorhabdi
24	89	10.0	287	16	Q89I38	Q89i38 bradyrhizob
25	87.5	9.8	1257	16	Q7WF24	Q7wf24 bordetella
26	87.5	9.8	1257	16	Q7W3P8	Q7w3p8 bordetella
27	87.5	9.8	1257	16	Q7VTE8	Q7vte8 bordetella
28	85	9.5	478	4	Q8IVV1	Q8ivv1 homo sapien
29	85	9.5	1649	4	Q8IWY7	Q8iwy7 homo sapien
30	82.5	9.2	208	16	Q92K28	Q92k28 rhizobium m
31	82	9.2	498	16	Q9KXK2	Q9kxk2 streptomyce
32	80.5	9.0	7716	16	Q7UWZ8	Q7uwz8 rhodopirell
33	80	9.0	409	16	Q82AG1	Q82ag1 streptomyce
34	80	9.0	493	16	Q98AT9	Q98at9 rhizobium l
35	80	9.0	651	5	Q9VFA9	Q9vfa9 drosophila
36	80	9.0	1024	16	Q8FRH1	Q8frh1 corynebacte
37	79.5	8.9	240	16	Q8F863	Q8f863 leptospira
38	79.5	8.9	716	10	Q8LJG8	Q8ljg8 oryza sativ
39	79	8.8	477	10	Q9SMY7	Q9smy7 arabidopsis
40	79	8.8	525	10	Q94JL8	Q94jl8 arabidopsis
41	78.5	8.8	285	16	Q98EU5	Q98eu5 rhizobium l
42	78	8.7	1715	6	Q9GLM4	Q9glm4 bos taurus
43	77.5	8.7	395	11	Q9QUP4	Q9qup4 mus musculu
44	77	8.6	263	5	Q7YZU8	Q7yzu8 hexamita in
45	77	8.6	314	16	Q8U8N8	Q8u8n8 agrobacteri

ALIGNMENTS

RESULT 1
Q8WXI5
ID Q8WXI5 PRELIMINARY; PRT; 187 AA.
AC Q8WXI5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Collagen XVIII (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21409408; PubMed=11517600;
RA Feng Y., Cui L.B., Liu C.X., Ma Q.J.;
RT "Inhibition effect in vitro of purified endostatin expressed in Pichia pastoris.";
RL Sheng Wu Gong Cheng Xue Bao 17:278-282(2001).
DR EMBL; AF416592; AAL37720.1; -.
FT NON_TER 1 1
SQ SEQUENCE 187 AA; 20448 MW; 72B1047D85838CD3 CRC64;

Query Match 100.0%; Score 893; DB 4; Length 187;
Best Local Similarity 100.0%; Pred. No. 2.7e-83;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 76

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         77 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 136

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 186
```

```
RESULT 2
Q8NG19
ID   Q8NG19      PRELIMINARY;      PRT;   261 AA.
AC   Q8NG19;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Multi-functional protein MFP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Dou D.;
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF282883; AAM52249.1; -.
SQ   SEQUENCE   261 AA;  26745 MW;  CA60C920AF3E90E5 CRC64;

  Query Match             100.0%;  Score 893;  DB 4;  Length 261;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-83;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         91 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 150

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        151 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 210

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        211 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 260


RESULT 3
Q8N4S4
ID   Q8N4S4      PRELIMINARY;      PRT;   816 AA.
AC   Q8N4S4;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Similar to collagen, type XVIII, alpha 1 (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Renal adenocarcinoma;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC033715; AAH33715.1; -.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   Pfam; PF01391; Collagen; 5.
DR   ProDom; PD000007; Clg_helix; 1.
KW   Collagen.
FT   NON_TER       1      1
SQ   SEQUENCE   816 AA;  82553 MW;  5D539B2946694F86 CRC64;

  Query Match             100.0%;  Score 893;  DB 4;  Length 816;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-82;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        646 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 705

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        706 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 765

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        766 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 815


RESULT 4
Q61434
ID   Q61434      PRELIMINARY;      PRT;  1140 AA.
AC   Q61434;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Collagen (Fragment).
GN   COL15A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Abe N., Muragaki Y., Yoshioka H., Inoue H., Ninomiya Y.;
RT   "Identification of a novel collagen chain represented by extensive
RT   interruptions in the triple-helical region.";
RL   Cell. Mol. Biol. Res. 196:576-582(1993).
DR   EMBL; D17546; BAA04483.1; -.
DR   PIR; B56101; B56101.
DR   HSSP; P39061; 1KOE.
DR   MGD; MGI:88449; Col15a1.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 8.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Clg_helix; 1.
KW   Collagen.
FT   NON_TER       1      1
SQ   SEQUENCE  1140 AA;  115156 MW;  8B0C7E6862B3BDFE CRC64;

  Query Match             87.1%;  Score 778;  DB 11;  Length 1140;
  Best Local Similarity   85.8%;  Pred. No. 1.5e-70;
  Matches  145;  Conservative   13;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||:|||||||||||||||||||||||:|||||||||||||||||||||| :||||
Db        969 VALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRGSVPIV 1028

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              |||||:| |||::|||||:| |:||||||||||:|||||| ||||||||||:|||| |
Db       1029 NLKDEVLSPSWDSLFSGSQGQLQPGARIFSFDGRDVLRHPAWPQKSVWHGSDPSGRRLME 1088

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 169
              |||||||||  |||||||| |||| | |||||::||||||||||||:
Db       1089 SYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMTS 1137


RESULT 5
Q9QZD2
ID   Q9QZD2      PRELIMINARY;      PRT;   226 AA.
AC   Q9QZD2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Collagen XVIII (Fragment).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=20227226; PubMed=10766159;
RA   Perletti G., Concari P., Giardini R., Marras E., Piccinini F.,
RA   Folkman J., Chen L.;
RT   "Antitumor activity of endostatin against carcinogen-induced rat
RT   primary mammary tumors.";
RL   Cancer Res. 60:1793-1796(2000).
DR   EMBL; AF189709; AAF00975.1; -.
DR   HSSP; P39061; 1KOE.
FT   NON_TER       1      1
SQ   SEQUENCE   226 AA;  25350 MW;  38B83C0486C0E949 CRC64;

  Query Match             85.4%;  Score 763;  DB 11;  Length 226;
  Best Local Similarity   84.6%;  Pred. No. 6.8e-70;
  Matches  143;  Conservative   12;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||:|||||||||||||||||||||||||:||||||||||||||||||||||||::||||
Db         55 VALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRSSVPIV 114

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              |||||:| |||: |||||:| |  |||||||||:|||||| ||||||||||:|||| |
Db        115 NLKDEVLSPSWDTLFSGSQGQLHSGARIFSFDGRDVLRHPAWPQKSVWHGSDPSGRRLME 174

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 169
              ||||||||||   |||||||| |||| | |  |||::||||||||||:
Db        175 SYCETWRTEATGVTGQASSLLSGRLLEQKAESCHNSYIVLCIENSFMTS 223


RESULT 6
Q9WUW5
ID   Q9WUW5      PRELIMINARY;      PRT;   171 AA.
AC   Q9WUW5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Collagen type XVIII, alpha (I) chain (Fragment).
GN   COL18A1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Jia J.D., Bauer M., Eberspaecher U., Donner P., Schuppan D.;
RT   "Partial 3'-terminal cDNA sequence of rat collagen XVIII/endostatin.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Jia J.D., Bauer M., Sedlaczek N., Ruehl M., Riecken E.O., Schuppan D.;
RT   "Temporospatial expression of collagen XVIII/endostatin in acute and
RT   chronic liver injuries.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ236873; CAB44263.1; -.
DR   HSSP; P39061; 1KOE.
FT   NON_TER       1      1
FT   NON_TER     171    171
SQ   SEQUENCE   171 AA;  18933 MW;  81BE2EE3FC2C8E72 CRC64;

  Query Match             84.8%;  Score 757;  DB 11;  Length 171;
  Best Local Similarity   85.0%;  Pred. No. 2e-69;
  Matches  142;  Conservative   11;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||:|||||||||||||||||||||||||:|||||||||||||||||||||||::||||
Db          5 VALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRSSVPIV 64

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              |||||:| |||: |||||:| |  |||||||||:|||||| ||||||||||:|||| |
Db         65 NLKDEVLSPSWDTLFSGSQGQLHSGARIFSFDGRDVLRHPAWPQKSVWHGSDPSGRRLME 124

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFM 167
              ||||||||||   |||||||| |||| | |  |||::|||||||||||
Db        125 SYCETWRTEATGVTGQASSLLSGRLLEQKAESCHNSYIVLCIENSFM 171


RESULT 7
Q9CRT2
ID   Q9CRT2      PRELIMINARY;      PRT;   160 AA.
AC   Q9CRT2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Procollagen, type XVIII, alpha 1 (Fragment).
GN   COL18A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryo;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK014292; BAB29249.1; -.
DR   HSSP; P39061; 1KOE.
DR   MGD; MGI:88451; Col18a1.
DR   GO; GO:0005604; C:basement membrane; IDA.
DR   GO; GO:0001525; P:angiogenesis; IMP.
FT   NON_TER       1      1
SQ   SEQUENCE   160 AA;  17725 MW;  60F853D777C375D2 CRC64;

  Query Match             80.9%;  Score 722;  DB 11;  Length 160;
  Best Local Similarity   85.4%;  Pred. No. 6.9e-66;
  Matches  134;  Conservative   12;  Mismatches   11;  Indels    0;  Gaps    0;

Qy         13 GIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIVNLKDELLFPSWE 72
              |||||||||||||||||:|||||||||||||||||||||| :|||||||||:| |||:
Db          1 GIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRGSVPIVNLKDEVLSPSWD 60

Qy         73 ALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTESYCETWRTEAPS 132
              :|||||:| |:||||||||||:|||||| ||||||||||:|||| ||||||||||
Db         61 SLFSGSQGQLQPGARIFSFDGRDVLRHPAWPQKSVWHGSDPSGRRLMESYCETWRTETTG 120

Qy        133 ATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 169
              ||||||||| |||| | |||||::||||||||||||:
Db        121 ATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMTS 157


RESULT 8
O93419
ID   O93419      PRELIMINARY;      PRT;  1344 AA.
AC   O93419;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
```

```
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Collagen XVIII precursor.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98411346; PubMed=9738008;
RA   Halfter W., Dong S., Schurer B., Cole G.J.;
RT   "Collagen XVIII is a basement membrane heparan sulfate proteoglycan.";
RL   J. Biol. Chem. 273:25404-25412(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Halfter W., Dong S.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF083440; AAC33294.2; -.
DR   HSSP; P39061; 1KOE.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 8.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Clg_helix; 2.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Collagen; Signal.
FT   SIGNAL       1     26       POTENTIAL.
SQ   SEQUENCE   1344 AA;  137402 MW;  7AA366E4FE940CCD CRC64;

  Query Match             79.6%;  Score 711;  DB 13;  Length 1344;
  Best Local Similarity   77.6%;  Pred. No. 1.4e-63;
  Matches  132;  Conservative   17;  Mismatches   21;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||:|||||||||||||||||||||| |||||||||||||||||||||||||||||| |||||
Db       1173 VALNTPLSGGMRGIRGADFQCFQQARQVGLAGTFRAFLSSRLQDLYSIVRRADRTAVPIV 1232

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||:||:|| :|||||:||| ||: |||| ||||:|:|:   |||||:|||||  |||| |
Db       1233 NLRDEVLFSNWEALFTGSEAPLRAGARILSFDGRDILQDSAWPQKSIWHGSDAKGRRLPE 1292

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              |||| |||:    :|||||  |:|| |||:|| ||::||||||||||:
Db       1293 SYCEAWRTDERGTSGQASSLSSGKLLEQSASSCQHAFVVLCIENSFMTAA 1342


RESULT 9
Q8QHL9
ID   Q8QHL9      PRELIMINARY;      PRT;  1315 AA.
AC   Q8QHL9;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Type XVIII collagen alpha1 chain.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ishino T., Sekimizu K., Natori S., Kubo T.;
RT   "Identification and characterization of genes expressed selectively in
RT   the regenerating tail of Xenopus laevis tadpole.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB047066; BAB84674.1; -.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 7.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Collagen.
SQ   SEQUENCE   1315 AA;  134946 MW;  0C56C235DE058365 CRC64;

  Query Match             73.2%;  Score 654;  DB 13;  Length 1315;
  Best Local Similarity   73.2%;  Pred. No. 9e-58;
  Matches  123;  Conservative   16;  Mismatches   29;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||:|||| |: ||| |||||:|||  || ||||||||||||||||||||||| :| ||
Db       1144 VALNAPLSGSMKSIRGVDFQCFEQARKSGLHGTFRAFLSSRLQDLYSIVRRADRQSVQIV 1203

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||:||:|: :||:||||||  ::|||||||||||||   |||||| ||||||  ||||||
Db       1204 NLRDEVLYDNWESLFSGSEAQMRPGARIFSFDGKDVATDPTWPQKMVWHGSDAKGRRLTE 1263

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 168
              ||||||||:  : | |||||  |:|| |   ||:  :|||||||||||
Db       1264 SYCETWRTDESAVTSQASSLTSGKLLEQRPQSCNKNFIVLCIENSFMT 1311


RESULT 10
Q8JFF7
ID   Q8JFF7      PRELIMINARY;      PRT;  1307 AA.
AC   Q8JFF7;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Type XVIII collagen short variant.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22166979; PubMed=12175494;
RA   Elamaa H., Peterson J., Pihlajaniemi T., Destree O.;
RT   "Cloning of three variants of type XVIII collagen and their expression
RT   patterns during Xenopus laevis development.";
RL   Mech. Dev. 114:109-113(2002).
DR   EMBL; AY052763; AAL14257.1; -.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 6.
DR   Pfam; PF02210; TSPN; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Collagen.
SQ   SEQUENCE   1307 AA;  134153 MW;  D53EDBFE3DE34976 CRC64;

  Query Match             72.0%;  Score 643;  DB 13;  Length 1307;
  Best Local Similarity   71.4%;  Pred. No. 1.2e-56;
  Matches  120;  Conservative   18;  Mismatches   30;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||:|:|| |: ||| |||||:|||  || ||||||||||||||||||||| :| ||
Db       1136 VALNAPISGSMKSIRGVDFQCFEQARKAGLHGTFRAFLSSRLQDLYSIVRRADRQSVQIV 1195

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||:||:|: :|::|||||  :: |||| |||||||   |||||| |||||  ||||||
Db       1196 NLRDEVLYDNWDSLFSGSEAQMRSGARILSFDGKDVTTDPTWPQKMVWHGSDAKGRRLTE 1255

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 168
              ||||||||:  : |||||||  |:|| |   ||:  :||||||||||
```

```
Db       1256 SYCETWRTDESAVTGQASSLTSGKLLEQRPQSCNKNFIVLCIENSFMT 1303


RESULT 11
Q8AWC6
ID   Q8AWC6      PRELIMINARY;      PRT;   361 AA.
AC   Q8AWC6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Collagen XVIII (Fragment).
GN   COL18A1.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Haftek Z., Morvan-Dubois G., Thisse B., Garrone R., Le Guellec D.;
RT   "Sequence and embryonic expression of collagen XVIII NC11 domain
RT   (endostatin) in the zebrafish.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ494837; CAD38825.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   361 AA;  40222 MW;  3C5A0F8479D26735 CRC64;

  Query Match             67.9%;  Score 606;  DB 13;  Length 361;
  Best Local Similarity   66.5%;  Pred. No. 1.4e-53;
  Matches  113;  Conservative   22;  Mismatches   35;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              :|||||  | |||||||||| |||||||||: |||||||||:||||||||||:||  :|||
Db        190 IALNSPQVGNMRGIRGADFMCFQQARAVGMKGTFRAFLSSKLQDLYSIVRRSDRETLPIV 249

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||::|| |||:||| ||  :|  | |:||||:||||    ||:| :|||||  | | |:
Db        250 NLKDQVLFSSWESLFSDSESRMKDNAPIYSFDGRDVLRDSAWPEKMIWHGSDGRGHRQTD 309

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              :||||||    : || ||||  |:|| |:::||  :|| ||||||:|| |
Db        310 NYCETWRAGDRAVTGLASSLQAGQLLQQTSSSCSSSYIALCIENSYMTQS 359


RESULT 12
Q96T70
ID   Q96T70      PRELIMINARY;      PRT;   102 AA.
AC   Q96T70;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Endostatin variant (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Deininger M.H., Trautmann K., Schluesener H.J.;
RT   "Endostatin promotes delayed secondary damage following traumatic
RT   brain injury.";
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF333247; AAK50626.1; -.
FT   NON_TER       1       1
FT   NON_TER     102     102
SQ   SEQUENCE   102 AA;  11147 MW;  ECAC47AA6420947D CRC64;

  Query Match             59.2%;  Score 529;  DB 4;  Length 102;
  Best Local Similarity   98.0%;  Pred. No. 2.2e-46;
  Matches  100;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         59 IVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRL 118
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||

Db          1 IVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTRPQKSVWHGSDPNGRRL 60

Qy        119 TESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVL 160
              |||||||||||||||||||||||||||||||||||||||||| |||||||
Db         61 TESYCETWRTEAPSATGQASSLLGGRLLGQSAASSHHAYIVL 102


RESULT 13
O35206
ID   O35206      PRELIMINARY;      PRT;  1367 AA.
AC   O35206;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type XV collagen.
GN   COL15A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97480713; PubMed=9339358;
RA   Hagg P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,
RA   Pihlajaniemi T.;
RT   "Cloning of mouse type XV collagen sequences and mapping of the
RT   corresponding gene to 4B1-3. Comparison of mouse and human alpha 1
RT   (XV) collagen sequences indicates divergence in the number of small
RT   collagenous domains.";
RL   Genomics 45:31-41(1997).
DR   EMBL; AF011450; AAC53387.1; -.
DR   PDB; 1DY2; 21-JAN-01.
DR   MGD; MGI:88449; Col15a1.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 5.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Collagen.
SQ   SEQUENCE   1367 AA;  140525 MW;  A483A1254AF3AEEC CRC64;

  Query Match             57.7%;  Score 515;  DB 11;  Length 1367;
  Best Local Similarity   58.3%;  Pred. No. 1.6e-43;
  Matches   98;  Conservative   26;  Mismatches   40;  Indels    4;  Gaps    1;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||:|::| :|    |||||||||| ||  |||||||| |||| ::||:|:|  :|||
Db       1200 VALNTPVAGDIR----ADFQCFQQARAAGLLSTFRAFLSSHLQDLSTVVRKAERFGLPIV 1255

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||| ::|| :|:::|||  |       |:||||:||:  |:|||| |||||:|:| || :
Db       1256 NLKGQVLFNNWDSIFSGDGGQFNTHIPIYSFDGRDVMTDPSWPQKVVWHGSNPHGVRLVD 1315

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 168
              ||| |||   : || || |  |::| | | || :  ||||||||||
Db       1316 KYCEAWRTTDMAVTGFASPLSTGKILDQKAYSCANRLIVLCIENSFMT 1363


RESULT 14
Q9EQD9
ID   Q9EQD9      PRELIMINARY;      PRT;  1367 AA.
AC   Q9EQD9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type XV collagen.
```

```
GN   COL15A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Sv;
RX   MEDLINE=97480713; PubMed=9339358;
RA   Hagg P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,
RA   Pihlajaniemi T.;
RT   "Cloning of mouse type XV collagen sequences and mapping of the
RT   corresponding gene to 4B1-3. Comparison of mouse and human alpha 1
RT   (XV) collagen sequences indicates divergence in the number of small
RT   collagenous domains.";
RL   Genomics 45:31-41(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Sv;
RX   MEDLINE=20522048; PubMed=11068203;
RA   Eklund L., Muona A., Lietard J., Pihlajaniemi T.;
RT   "Structure of the mouse type XV collagen gene, Col15a1, comparison
RT   with the human COL15A1 gene and functional analysis of the promoters
RT   of both genes.";
RL   Matrix Biol. 19:489-500(2000).
DR   EMBL; AF261131; AAG27545.1; -.
DR   EMBL; AF261109; AAG27545.1; JOINED.
DR   EMBL; AF261110; AAG27545.1; JOINED.
DR   EMBL; AF261111; AAG27545.1; JOINED.
DR   EMBL; AF261112; AAG27545.1; JOINED.
DR   EMBL; AF261113; AAG27545.1; JOINED.
DR   EMBL; AF261114; AAG27545.1; JOINED.
DR   EMBL; AF261115; AAG27545.1; JOINED.
DR   EMBL; AF261116; AAG27545.1; JOINED.
DR   EMBL; AF261117; AAG27545.1; JOINED.
DR   EMBL; AF261118; AAG27545.1; JOINED.
DR   EMBL; AF261119; AAG27545.1; JOINED.
DR   EMBL; AF261120; AAG27545.1; JOINED.
DR   EMBL; AF261121; AAG27545.1; JOINED.
DR   EMBL; AF261122; AAG27545.1; JOINED.
DR   EMBL; AF261123; AAG27545.1; JOINED.
DR   EMBL; AF261124; AAG27545.1; JOINED.
DR   EMBL; AF261125; AAG27545.1; JOINED.
DR   EMBL; AF261126; AAG27545.1; JOINED.
DR   EMBL; AF261127; AAG27545.1; JOINED.
DR   EMBL; AF261128; AAG27545.1; JOINED.
DR   EMBL; AF261129; AAG27545.1; JOINED.
DR   EMBL; AF261130; AAG27545.1; JOINED.
DR   HSSP; P39061; 1KOE.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 5.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Collagen.
SQ   SEQUENCE   1367 AA;  140454 MW;  40F259A3CD47C982 CRC64;

  Query Match             57.3%;  Score 512;  DB 11;  Length 1367;
  Best Local Similarity   58.3%;  Pred. No. 3.2e-43;
  Matches   98;  Conservative   25;  Mismatches   41;  Indels    4;  Gaps    1;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||:|::| :|    |||||||||| ||  |||||||| |||| ::||:|:|  :|||
Db       1200 VALNTPVAGDIR----ADFQCFQQARAAGLLSTFRAFLSSHLQDLSTVVRKAERFGLPIV 1255

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||| ::|| :|:::|||  |       |:||||:||:  |:|||| |||||:| | || :
Db       1256 NLKGQVLFNNWDSIFSGDGGQFNTHIPIYSFDGRDVMTDPSWPQKVVWHGSNPYGVRLVD 1315

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 168
               ||| |||   : || || |  |::| | | || :  ||||||||||
Db       1316 KYCEAWRTTDMAVTGFASPLSTGKILDQKAYSCANRLIVLCIENSFMT 1363


RESULT 15
Q9Y4W4
ID   Q9Y4W4      PRELIMINARY;      PRT;  1388 AA.
AC   Q9Y4W4;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type XV collagen.
GN   COL15A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94148920; PubMed=8106446;
RA   Kivirikko S., Heinamaki P., Rehn M., Honkanen N., Myers J.C.,
RA   Pihlajaniemi T.;
RT   "Primary structure of the alpha 1 chain of human type XV collagen and
RT   exon-intron organization in the 3' region of the corresponding gene.";
RL   J. Biol. Chem. 269:4773-4779(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98316357; PubMed=9651385;
RA   Hagg P.M., Muona A., Lietard J., Kivirikko S., Pihlajaniemi T.;
RT   "Complete exon-intron organization of the human gene for the alpha1
RT   chain of type XV collagen (COL15A1) and comparison with the homologous
RT   COL18A1 gene.";
RL   J. Biol. Chem. 273:17824-17831(1998).
DR   EMBL; L25280; AAC78500.1; -.
DR   EMBL; AF052956; AAC78500.1; JOINED.
DR   EMBL; AF052957; AAC78500.1; JOINED.
DR   EMBL; AF052958; AAC78500.1; JOINED.
DR   EMBL; AF052959; AAC78500.1; JOINED.
DR   EMBL; AF052960; AAC78500.1; JOINED.
DR   EMBL; AF052961; AAC78500.1; JOINED.
DR   EMBL; AF052962; AAC78500.1; JOINED.
DR   EMBL; AF052963; AAC78500.1; JOINED.
DR   EMBL; AF052964; AAC78500.1; JOINED.
DR   EMBL; AF052965; AAC78500.1; JOINED.
DR   EMBL; AF052966; AAC78500.1; JOINED.
DR   EMBL; AF052967; AAC78500.1; JOINED.
DR   EMBL; AF052968; AAC78500.1; JOINED.
DR   EMBL; AF052969; AAC78500.1; JOINED.
DR   EMBL; AF052970; AAC78500.1; JOINED.
DR   EMBL; AF052971; AAC78500.1; JOINED.
DR   EMBL; AF052972; AAC78500.1; JOINED.
DR   EMBL; AF052973; AAC78500.1; JOINED.
DR   EMBL; AF052974; AAC78500.1; JOINED.
DR   EMBL; AF052975; AAC78500.1; JOINED.
DR   EMBL; L25285; AAC78500.1; JOINED.
DR   EMBL; L25284; AAC78500.1; JOINED.
DR   EMBL; L25283; AAC78500.1; JOINED.
DR   EMBL; L25282; AAC78500.1; JOINED.
DR   EMBL; L25281; AAC78500.1; JOINED.
DR   HSSP; P39061; 1KOE.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 5.
DR   Pfam; PF02210; TSPN; 1.
DR   SMART; SM00282; LamG; 1.
```

```
DR   SMART; SM00210; TSPN; 1.
KW   Collagen.
SQ   SEQUENCE   1388 AA;  141757 MW;  96828E45E847194B CRC64;

  Query Match             56.6%;  Score 505;  DB 4;  Length 1388;
  Best Local Similarity   56.9%;  Pred. No. 1.7e-42;
  Matches   95;  Conservative   28;  Mismatches   40;  Indels    4;  Gaps    1;

Qy          2 ALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIVN 61
              ||| |:|| :|    |||||||:|||| ||  |:|||||| |||| :|||:|:| ::|||| 
Db       1222 ALNMPVSGDIR----ADFQCFKQARAAGLLSTYRAFLSSHLQDLSTIVRKAERYSLPIVN 1277

Qy         62 LKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTES 121
              || ::|| :|:::||| |       |:||||:|::  |:|||| :|||| |:| || ::
Db       1278 LKGQVLFNNWDSIFSGHGGQFNMHIPIYSFDGRDIMTDPSWPQKVIWHGSSPHGVRLVDN 1337

Qy        122 YCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 168
              ||| |||   : || || |  |::| | | || :  ||||||||||
Db       1338 YCEAWRTADTAVTGLASPLSTGKILDQKAYSCANRLIVLCIENSFMT 1384


Search completed: March 13, 2004, 08:17:54
Job time : 44 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 13, 2004, 08:27:58 ; Search time 58 Seconds
(without alignments)
828.156 Million cell updates/sec

Title: US-09-171-607A-1
Perfect score: 893
Sequence: 1 VALNSPLSGGMRGIRGADFQ.....ASCHHAYIVLCIENSFMTAS 170

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	893	100.0	178	3	AAY94324	Aay94324 Alternate
2	893	100.0	178	4	AAU00900	Aau00900 Human End
3	893	100.0	179	4	AAU00901	Aau00901 Human End
4	893	100.0	182	3	AAY59622	Aay59622 Human end
5	893	100.0	182	3	AAY94323	Aay94323 Human end
6	893	100.0	182	3	AAB28399	Aab28399 Human end
7	893	100.0	182	4	AAU00897	Aau00897 Human End
8	893	100.0	182	5	AAU77951	Aau77951 Amino aci
9	893	100.0	183	2	AAY02113	Aay02113 SEQ ID 76
10	893	100.0	183	2	AAY08693	Aay08693 Human end
11	893	100.0	183	3	AAY70252	Aay70252 Human ang
12	893	100.0	183	3	AAY90771	Aay90771 Human ang
13	893	100.0	183	3	AAB16451	Aab16451 Human end
14	893	100.0	183	3	AAB30493	Aab30493 Amino aci
15	893	100.0	183	4	AAB49379	Aab49379 Human end
16	893	100.0	183	4	AAU00896	Aau00896 Human End
17	893	100.0	183	5	ABB79901	Abb79901 Human end
18	893	100.0	183	5	AAM49503	Aam49503 Human end
19	893	100.0	183	5	AAM48895	Aam48895 Human end
20	893	100.0	183	5	AAU97132	Aau97132 Human end
21	893	100.0	183	6	AAG79753	Aag79753 Human end
22	893	100.0	193	5	AAU76690	Aau76690 Synthetic
23	893	100.0	195	3	AAW90874	Aaw90874 Human HMW
24	893	100.0	216	3	AAB30495	Aab30495 Amino aci
25	893	100.0	275	5	AAU76689	Aau76689 Synthetic
26	893	100.0	310	5	AAU76688	Aau76688 Human col
27	893	100.0	513	5	ABG73586	Abg73586 Human End
28	893	100.0	682	5	ABP41878	Abp41878 Human ova
29	893	100.0	684	2	AAW26327	Aaw26327 Human alp
30	893	100.0	684	2	AAY25113	Aay25113 Human alp
31	893	100.0	684	5	AAO17357	Aao17357 Human col
32	893	100.0	1301	2	AAW92296	Aaw92296 Human alp
33	893	100.0	1336	2	AAY08694	Aay08694 Human col
34	893	100.0	1336	6	ABP96308	Abp96308 Human end
35	893	100.0	1516	5	ABB83471	Abb83471 Human col
36	893	100.0	1516	5	ABP68617	Abp68617 Human pan
37	889	99.6	181	4	AAU00898	Aau00898 Human End
38	889	99.6	183	4	AAB49810	Aab49810 Human end
39	889	99.6	184	5	ABG31794	Abg31794 Murine en
40	886	99.2	184	4	AAG78717	Aag78717 Human vas
41	885	99.1	180	4	AAU00899	Aau00899 Human End
42	885	99.1	193	3	AAW90877	Aaw90877 Human HMW
43	882	98.8	271	3	AAB08407	Aab08407 A human a
44	876	98.1	184	7	ADE40256	Ade40256 Human NOV
45	783	87.7	184	3	AAY70265	Aay70265 Canine an

ALIGNMENTS

RESULT 1
AAY94324
ID AAY94324 standard; protein; 178 AA.
XX
AC AAY94324;
XX
DT 11-AUG-2000 (first entry)
XX
DE Alternate human endostatin protein.
XX
KW Human; endothelial cell proliferation inhibitor; collagen XVIII;
KW angiogenesis inhibitor; anti-tumour; cytostatic; antipsoriatic;
KW vasotropic; determatological; ophthalmological; vulnerary;
KW antiarteriosclerotic; antidiabetic; haemostatic; contraceptive;
KW ocular angiogenic disease; atherosclerosis; scleroderma;
KW myocardial angiogenesis; telangiectasia; angiofibroma; wound granulation.
XX
OS Homo sapiens.
XX
PN WO200026368-A2.
XX
PD 11-MAY-2000.
XX
PF 01-NOV-1999; 99WO-US025605.
XX
PR 30-OCT-1998; 98US-0106343P.
PR 20-MAY-1999; 99US-00315689.
XX
PA (CHIL-) CHILDRENS MEDICAL CENT.
XX
PI O'reilly MS, Folkman MJ;
XX
DR WPI; 2000-365617/31.
DR N-PSDB; AAA27005.
XX
PT Novel endostatin capable of inhibiting endothelial cell proliferation and
PT angiogenesis, useful for treating angiogenesis-dependent cancers and as
PT birth control agents.
XX
PS Claim 3; Page 39; 68pp; English.
XX
CC The present sequence is an alternate functional endostatin protein. When
CC the human endostatin gene sequence AAA27004 is recombinantly expressed,
CC an observable doublet of protein results, both versions of which are
CC functional endostatin proteins. The present endostatin variant is the
CC same as the protein encoded by AAA27004 minus the first four amino acids.
CC Recombinant mouse endostatin (20 mg/kg) was administered subcutaneously
CC to mice implanted with Lewis lung carcinomas. There was tumour mass

CC regression non-detectable levels after 12 days of therapy due to the
CC angiogenesis inhibitory activity of endostatin. Thus the protein is
CC useful for treatment of angiogenesis- dependent cancers. The
CC polynucleotide and polypeptide sequences of this endostatin are useful
CC for treating and diagnosis of tumours, ocular angiogenic diseases, Osler-
CC Webber syndrome, myocardial angiogenesis, plaque neovascularisation,
CC telangiectasia, haemophiliac joints, angiofibroma and wound granulation,
CC for treatment of diseases related to excessive or abnormal stimulation of
CC endothelial cells e.g. intestinal adhesions, atherosclerosis,
CC scleroderma. The protein may also be useful as a birth control agent by
CC reducing or preventing uterine vascularisation. The gene for endostatin
CC may be isolated from cells or tissue that express high levels of
CC endostatin, eg. tumour cells, by generating cDNA from mRNA using reverse
CC transcriptase and then amplifying the DNA sequence
XX
SQ Sequence 178 AA;

Query Match 100.0%; Score 893; DB 3; Length 178;
Best Local Similarity 100.0%; Pred. No. 3.1e-101;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          9 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 68

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 128

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 178
```

RESULT 2
AAU00900
ID AAU00900 standard; protein; 178 AA.
XX
AC AAU00900;
XX
DT 04-JUL-2001 (first entry)
XX
DE Human Endostatin(TM) N-terminal deletion mutant protein#2.
XX
KW Human; Endostatin(TM); angiogenesis mediated disease; solid tumours;
KW blood borne tumour; leukaemia; tumour metastasis; benign tumour;
KW haemangioma; acoustic neuroma; neurofibroma; trachoma; rubeosis;
KW pyogenic granuloma; rheumatoid arthritis; psoriasis; colon cancer;
KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW retinopathy of prematurity; macular corneal graft rejection;
KW neovascular glaucoma; retrolental fibroplasia; Osler-Webber Syndrome;
KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW haemophiliac joint; angiofibroma; wound granulation; variant; mutant;
KW mutein.
XX
OS Homo sapiens.
XX
PN WO200119989-A2.
XX
PD 22-MAR-2001.
XX
PF 14-SEP-2000; 2000WO-US025166.
XX
PR 14-SEP-1999; 99US-0153698P.
XX
PA (ENTR-) ENTREMED INC.
XX
PI Liang H, Sim KL, Chang-Murad A, Zhou X, Madsen J, Boerner RJ;
PI Bermejo LL, Mistry FR, Shepard SR, Schrimsher JL;
XX
DR WPI; 2001-244802/25.
DR N-PSDB; AAS00868.
XX
PT Producing Endostatin protein for treating angiogenesis mediated diseases
PT such as solid tumors, comprises recombinantly producing the protein using
PT an expression system, and recovering and purifying the protein.
XX
PS Claim 5; Page 33; 67pp; English.
XX
CC The sequence represents Human Endostatin(TM) N-terminal deletion mutant
CC protein lacking the N-terminal 4 amino acids and the C-terminal lysine, a
CC natural variant recovered from fermentations of Pichia pastoris cultures
CC harbouring a expression plasmid containing the Endostatin(TM) DNA
CC sequence given in AAS00868. The new method of the invention is useful for
CC producing, recovering and purifying Endostatin (TM) from biological
CC sources, such as biological fluids, tissues, cells, culture media, and
CC fermentation media. Endostatin(TM) is useful for treating angiogenesis
CC mediated diseases such as solid tumours, blood borne tumours, leukaemias,
CC tumour metastases, benign tumours, e.g. haemangioma, acoustic neuromas,
CC neurofibromas, trachomas, and pyogenic granulomas, rheumatoid arthritis,
CC psoriasis, ocular angiogenic diseases, e.g., diabetic retinopathy,
CC retinopathy of prematurity, macular degeneration, corneal graft
CC rejection, neovascular glaucoma, colon cancer, retrolental fibroplasia,
CC rubeosis, Osler-Webber Syndrome, myocardial angiogenesis, plaque
CC neovascularisation, telangiectasia, haemophiliac joints, angiofibroma,
CC and wound granulation. Endostatin(TM) is also useful for treating disease
CC of excessive or abnormal stimulation of endothelial cells such as
CC intestinal adhesions, atherosclerosis, scleroderma and hypertrophic
CC scars. Higher yields of more purified, and biologically active
CC Endostatin(TM) are obtained by the new method. Endostatin(TM) can be
CC stored in buffers for extended periods of time, and also subjected to
CC lyophilisation, while preserving biological activity. Centrifugation of
CC broth from fermentation steps in production is avoided, preventing
CC unwanted potential cellular lysis and contamination with additional
CC proteins, pigments, enzymes and other cellular chemicals and debris
XX
SQ Sequence 178 AA;

Query Match 100.0%; Score 893; DB 4; Length 178;
Best Local Similarity 100.0%; Pred. No. 3.1e-101;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          9 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 68

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 128

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 178
```

RESULT 3
AAU00901
ID AAU00901 standard; protein; 179 AA.
XX
AC AAU00901;
XX
DT 04-JUL-2001 (first entry)
XX
DE Human Endostatin(TM) N-terminal mutant protein#1.
XX
KW Human; Endostatin(TM); angiogenesis mediated disease; solid tumours;
KW blood borne tumour; leukaemia; tumour metastasis; benign tumour;
KW haemangioma; acoustic neuroma; neurofibroma; trachoma; rubeosis;
KW pyogenic granuloma; rheumatoid arthritis; psoriasis; colon cancer;
KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW retinopathy of prematurity; macular corneal graft rejection;
KW neovascular glaucoma; retrolental fibroplasia; Osler-Webber Syndrome;
KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW haemophiliac joint; angiofibroma; wound granulation; mutant; mutein.
XX
OS Homo sapiens.

XX
PN WO200119989-A2.
XX
PD 22-MAR-2001.
XX
PF 14-SEP-2000; 2000WO-US025166.
XX
PR 14-SEP-1999; 99US-0153698P.
XX
PA (ENTR-) ENTREMED INC.
XX
PI Liang H, Sim KL, Chang-Murad A, Zhou X, Madsen J, Boerner RJ;
PI Bermejo LL, Mistry FR, Shepard SR, Schrimsher JL;
XX
DR WPI; 2001-244802/25.
DR N-PSDB; AAS00868.
XX
PT Producing Endostatin protein for treating angiogenesis mediated diseases
PT such as solid tumors, comprises recombinantly producing the protein using
PT an expression system, and recovering and purifying the protein.
XX
PS Claim 5; Page 32; 67pp; English.
XX
CC The sequence represents a Human Endostatin(TM) N-terminal deletion mutant
CC lacking the N-terminal 4 amino acids. The new method of the invention is
CC useful for producing, recovering and purifying Endostatin (TM) from
CC biological sources, such as biological fluids, tissues, cells, culture
CC media, and fermentation media. Endostatin(TM) is useful for treating
CC angiogenesis mediated diseases such as solid tumours, blood borne
CC tumours, leukaemias, tumour metastases, benign tumours, e.g. haemangioma,
CC acoustic neuromas, neurofibromas, trachomas, and pyogenic granulomas,
CC rheumatoid arthritis, psoriasis, ocular angiogenic diseases, e.g.,
CC diabetic retinopathy, retinopathy of prematurity, macular degeneration,
CC corneal graft rejection, neovascular glaucoma, colon cancer, retrolental
CC fibroplasia, rubeosis, Osler-Webber Syndrome, myocardial angiogenesis,
CC plaque neovascularisation, telangiectasia, haemophiliac joints,
CC angiofibroma, and wound granulation. Endostatin(TM) is also useful for
CC treating disease of excessive or abnormal stimulation of endothelial
CC cells such as intestinal adhesions, atherosclerosis, scleroderma and
CC hypertrophic scars. Higher yields of more purified, and biologically
CC active Endostatin(TM) are obtained by the new method. Endostatin(TM) can
CC be stored in buffers for extended periods of time, and also subjected to
CC lyophilisation, while preserving biological activity. Centrifugation of
CC broth from fermentation steps in production is avoided, preventing
CC unwanted potential cellular lysis and contamination with additional
CC proteins, pigments, enzymes and other cellular chemicals and debris
XX
SQ Sequence 179 AA;

Query Match 100.0%; Score 893; DB 4; Length 179;
Best Local Similarity 100.0%; Pred. No. 3.1e-101;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          9 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 68

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 128

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 178
```

RESULT 4
AAY59622
ID AAY59622 standard; protein; 182 AA.
XX
AC AAY59622;
XX
DT 14-MAR-2000 (first entry)

XX
DE Human endostatin protein fragment.
XX
KW Endostatin; scatter factor activity; human; tubulogenesis; psoriasis;
KW metastatic cancer; tumourigenesis; ocular angiogenic disease;
KW rheumatoid arthritis; Osler-Webber syndrome; telangiectasia;
KW haemophiliac joint; angiofibroma; wound granulation.
XX
OS Homo sapiens.
XX
PN WO9962944-A2.
XX
PD 09-DEC-1999.
XX
PF 03-JUN-1999; 99WO-US012278.
XX
PR 03-JUN-1998; 98US-0087890P.
PR 10-JUL-1998; 98US-0092393P.
PR 01-SEP-1998; 98US-0098790P.
XX
PA (CHIL-) CHILDRENS MEDICAL CENT.
XX
PI Javaherian K, Folkman MJ;
XX
DR WPI; 2000-072833/06.
XX
PT New endostatin oligomers, used for treating e.g. tumors.
XX
PS Disclosure; Page 6; 44pp; English.
XX
CC This sequence is a fragment of the human endostatin protein. Endostatin
CC is an approximately 20kD C-terminal globular domain of the collagen-like
CC protein collagen XVIII. Protein oligomers consisting of more than one
CC endostatin monomer have anti-tubulogenic effects and induce
CC reorganization of the actin cytoskeleton. The oligomer has scatter factor
CC activity. The oligomers induce the destruction of tubular lumens and
CC elongation of cells, and inhibit tubulogenesis and tumourigenesis. The
CC oligomers can also be used to treat metastatic cancers, tumours,
CC rheumatoid arthritis, psoriasis, ocular angiogenic disease, Osler-Webber
CC syndrome, plaque neovascularisation, telangiectasia, haemophiliac joints,
CC angiofibroma and wound granulation. The oligomers can also be used to
CC treat diseases that have angiogenesis as a pathological consequence e.g.
CC ulcers. The endostatin oligomers can also be used to develop affinity
CC columns for isolating antibodies or receptors. Passive antibody therapy
CC using antibodies that specifically bind endostatin oligomers can be used
CC to modulate morphogenic processes such as metastatic cancer as well as
CC angiogenesis-dependent processes such as reproduction, development, wound
CC healing, tissue repair, and angiogenesis-dependent diseases. Also,
CC antisera directed to the Fab regions of endostatin oligomer antibodies
CC can be administered to block the ability of endogenous endostatin
CC oligomer antisera to bind endostatin oligomers
XX
SQ Sequence 182 AA;

Query Match 100.0%; Score 893; DB 3; Length 182;
Best Local Similarity 100.0%; Pred. No. 3.2e-101;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182
```

RESULT 5
AAY94323

```
ID   AAY94323 standard; protein; 182 AA.
XX
AC   AAY94323;
XX
DT   11-AUG-2000  (first entry)
XX
DE   Human endostatin protein.
XX
KW   Human; endothelial cell proliferation inhibitor; collagen XVIII;
KW   angiogenesis inhibitor; anti-tumour; cytostatic; antipsoriatic;
KW   vasotropic; determatological; ophthalmological; vulnerary;
KW   antiarteriosclerotic; antidiabetic; haemostatic; contraceptive;
KW   ocular angiogenic disease; atherosclerosis; scleroderma;
KW   myocardial angiogenesis; telangiectasia; angiofibroma; wound granulation.
XX
OS   Homo sapiens.
XX
PN   WO200026368-A2.
XX
PD   11-MAY-2000.
XX
PF   01-NOV-1999;   99WO-US025605.
XX
PR   30-OCT-1998;   98US-0106343P.
PR   20-MAY-1999;   99US-00315689.
XX
PA   (CHIL-) CHILDRENS MEDICAL CENT.
XX
PI   O'reilly MS,  Folkman MJ;
XX
DR   WPI; 2000-365617/31.
DR   N-PSDB; AAA27004.
XX
PT   Novel endostatin capable of inhibiting endothelial cell proliferation and
PT   angiogenesis, useful for treating angiogenesis-dependent cancers and as
PT   birth control agents.
XX
PS   Claim 2; Page 38; 68pp; English.
XX
CC   The present sequence is an endostatin protein which is the carboxy
CC   terminal protein of human collagen XVIII. Recombinant mouse endostatin
CC   (20 mg/kg) was administered subcutaneously to mice implanted with Lewis
CC   lung carcinomas. There was tumour mass regression non-detectable levels
CC   after 12 days of therapy due to the angiogenesis inhibitory activity of
CC   endostatin. Thus the protein is useful for treatment of angiogenesis-
CC   dependent cancers. The polynucleotide and polypeptide sequences of this
CC   endostatin are useful for treating and diagnosis of tumours, ocular
CC   angiogenic diseases, Osler-Webber syndrome, myocardial angiogenesis,
CC   plaque neovascularisation, telangiectasia, haemophiliac joints,
CC   angiofibroma and wound granulation, for treatment of diseases related to
CC   excessive or abnormal stimulation of endothelial cells e.g. intestinal
CC   adhesions, atherosclerosis, scleroderma. The protein may also be useful
CC   as a birth control agent by reducing or preventing uterine
CC   vascularisation. The gene for endostatin may be isolated from cells or
CC   tissue that express high levels of endostatin, eg. tumour cells, by
CC   generating cDNA from mRNA using reverse transcriptase and then amplifying
CC   the DNA sequence
XX
SQ   Sequence 182 AA;

  Query Match             100.0%;  Score 893;  DB 3;  Length 182;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 6
AAB28399
ID   AAB28399 standard; protein; 182 AA.
XX
AC   AAB28399;
XX
DT   19-FEB-2001  (first entry)
XX
DE   Human endostatin.
XX
KW   Human; endostatin; cytostatic; antiproliferative;
KW   vascular endothelial growth factor; VEGF; antibody; VEGF2 receptor;
KW   cancer; vascularised solid tumour.
XX
OS   Homo sapiens.
XX
PN   WO200064946-A2.
XX
PD   02-NOV-2000.
XX
PF   28-APR-2000; 2000WO-US011367.
XX
PR   28-APR-1999;   99US-0131432P.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Thorpe PE,  Brekken RA;
XX
DR   WPI; 2000-687317/67.
XX
PT   Immunogenic composition for the treatment and diagnosis of cancer
PT   comprises an anti-VEGF (vascular endothelial growth factor) antibody
PT   binding the same epitope as the monoclonal antibody ATCC PTA 1595.
XX
PS   Example 10; Page 291-292; 298pp; English.
XX
CC   The present invention relates to anti-Vascular Endothelial Growth Factor
CC   (VEGF) antibodies that bind to the same epitope as the monoclonal
CC   antibody ATCC PTA 1595 and which significantly inhibit VEGF binding to
CC   the VEGF receptor VEGFR2, without inhibiting VEGF binding to the VEGF
CC   receptor VEGFR1. The present sequence is human endostatin. Endostatin may
CC   be conjugated onto the anti-VEGF antibodies of the present invention. The
CC   anti-VEGF antibodies of the present invention are useful for the
CC   treatment and diagnosis of cancer, especially vascularised solid tumours
XX
SQ   Sequence 182 AA;

  Query Match             100.0%;  Score 893;  DB 3;  Length 182;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 7
AAU00897
ID   AAU00897 standard; protein; 182 AA.
XX
AC   AAU00897;
```

XX
DT 04-JUL-2001 (first entry)
XX
DE Human Endostatin(TM) C-terminus minus 1 protein.
XX
KW Human; Endostatin(TM); angiogenesis mediated disease; solid tumours;
KW blood borne tumour; leukaemia; tumour metastasis; benign tumour;
KW haemangioma; acoustic neuroma; neurofibroma; trachoma; rubeosis;
KW pyogenic granuloma; rheumatoid arthritis; psoriasis; colon cancer;
KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW retinopathy of prematurity; macular corneal graft rejection;
KW neovascular glaucoma; retrolental fibroplasia; Osler-Webber Syndrome;
KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW haemophiliac joint; angiofibroma; wound granulation; variant;
KW C-terminus minus 1 protein.
XX
OS Homo sapiens.
XX
PN WO200119989-A2.
XX
PD 22-MAR-2001.
XX
PF 14-SEP-2000; 2000WO-US025166.
XX
PR 14-SEP-1999; 99US-0153698P.
XX
PA (ENTR-) ENTREMED INC.
XX
PI Liang H, Sim KL, Chang-Murad A, Zhou X, Madsen J, Boerner RJ;
PI Bermejo LL, Mistry FR, Shepard SR, Schrimsher JL;
XX
DR WPI; 2001-244802/25.
DR N-PSDB; AAS00897.
XX
PT Producing Endostatin protein for treating angiogenesis mediated diseases
PT such as solid tumors, comprises recombinantly producing the protein using
PT an expression system, and recovering and purifying the protein.
XX
PS Claim 5; Page 30; 67pp; English.
XX
CC The sequence represents Human Endostatin(TM) C-terminus minus 1 protein,
CC a natural variant lacking the C-terminal amino acid of Endostatin(TM)
CC recovered from fermentations of Pichia pastoris cultures harbouring a
CC expression plasmid containing the Endostatin(TM) DNA sequence given in
CC AAS00867. The new method of the invention is useful for producing,
CC recovering and purifying Endostatin (TM) from biological sources, such as
CC biological fluids, tissues, cells, culture media, and fermentation media.
CC Endostatin(TM) is useful for treating angiogenesis mediated diseases such
CC as solid tumours, blood borne tumours, leukaemias, tumour metastases,
CC benign tumours, e.g. haemangioma, acoustic neuromas, neurofibromas,
CC trachomas, and pyogenic granulomas, rheumatoid arthritis, psoriasis,
CC ocular angiogenic diseases, e.g., diabetic retinopathy, retinopathy of
CC prematurity, macular degeneration, corneal graft rejection, neovascular
CC glaucoma, colon cancer, retrolental fibroplasia, rubeosis, Osler-Webber
CC Syndrome, myocardial angiogenesis, plaque neovascularisation,
CC telangiectasia, haemophiliac joints, angiofibroma, and wound granulation.
CC Endostatin(TM) is also useful for treating disease of excessive or
CC abnormal stimulation of endothelial cells such as intestinal adhesions,
CC atherosclerosis, scleroderma and hypertrophic scars. Higher yields of
CC more purified, and biologically active Endostatin(TM) are obtained by the
CC new method. Endostatin(TM) can be stored in buffers for extended periods
CC of time, and also subjected to lyophilisation, while preserving
CC biological activity. Centrifugation of broth from fermentation steps in
CC production is avoided, preventing unwanted potential cellular lysis and
CC contamination with additional proteins, pigments, enzymes and other
CC cellular chemicals and debris
XX
SQ Sequence 182 AA;

Query Match 100.0%; Score 893; DB 4; Length 182;
Best Local Similarity 100.0%; Pred. No. 3.2e-101;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182
```

RESULT 8
AAU77951
ID AAU77951 standard; protein; 182 AA.
XX
AC AAU77951;
XX
DT 02-JUL-2002 (first entry)
XX
DE Amino acid sequence for human endostatin.
XX
KW Human; immunoconjugate; anti-vascular endothelial growth factor antibody;
KW anti-VEGF antibody; monoclonal antibody 2C3 ATCC PTA 1595; VEGF receptor;
KW VEGFR2; KDR/Flk-1; VEGFR1; Flt-1; angiogenesis; macular degeneration;
KW ocular neovascular disease; cancer; vascularised solid tumour; AIDS;
KW metastatic tumour; endothelial cell proliferation; inflammatory disorder;
KW atherosclerosis; diabetic retinopathy; corneal graft rejection;
KW acquired immune deficiency syndrome; infection; restenosis; fungal ulcer;
KW sickle cell anaemia; endometriosis; endostatin.
XX
OS Homo sapiens.
XX
PN AU200179401-A.
XX
PD 06-DEC-2001.
XX
PF 12-OCT-2001; 2001AU-00079401.
XX
PR 28-APR-2000; 2000AU-00048049.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Thorpe PE, Brekken RA;
XX
DR WPI; 2002-281368/33.
XX
PT Immunoconjugate compositions for treating cancer by inhibiting
PT angiogenesis and for delivering a diagnostic agent to tumor, comprises
PT anti-vascular endothelial growth factor antibody attached to a biological
PT agent.
XX
PS Example 10; Page 12-13 (Sequence listing); 299pp; English.
XX
CC The present invention relates to antibody-based compositions comprising
CC an immunoconjugate such as anti-vascular endothelial growth factor (VEGF)
CC antibody (Ab) (or its antigen-binding fragment), attached to a biological
CC agent, where the Ab binds to the same epitope as the monoclonal antibody
CC (MAb) 2C3 ATCC PTA 1595, and significantly inhibits VEGF binding to the
CC VEGF receptor VEGFR2 (KDR/Flk-1) without inhibiting VEGF binding to the
CC VEGF receptor VEGFR1 (Flt-1). The compositions of the invention are
CC useful in therapy, and diagnosis, for inhibiting angiogenesis in an
CC animal having ocular neovascular disease or macular degeneration, and for
CC delivering a biological agent to a vascularised tumour. The compositions
CC can also be used for treating cancer and subjects at risk of developing,
CC a vascularised solid tumour, a metastatic tumour or metastases from a
CC primary tumour. The composition is useful for specifically inhibiting
CC VEGF-induced endothelial cell proliferation, without significantly
CC inhibiting VEGF-induced macrophage, osteoclast or chondroclast function.
CC The compositions can be used for treating various diseases such as
CC inflammatory disorders, atherosclerosis, diabetic retinopathy,
CC restenosis, acquired immune deficiency syndrome (AIDS), blood borne

```
CC   tumours, corneal graft rejection, Crohn's disease, fungal ulcers,
CC   infections, sickle cell anaemia, and endometriosis. The present sequence
CC   represents human endostatin. Endostatin may be attached or functionally
CC   associated with anti-VEGF antibodies
XX
SQ   Sequence 182 AA;

  Query Match             100.0%;  Score 893;  DB 5;  Length 182;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 9
AAY02113
ID   AAY02113 standard; protein; 183 AA.
XX
AC   AAY02113;
XX
DT   16-JUL-1999  (first entry)
XX
DE   SEQ ID 76 of WO9916889.
XX
KW   Angiostatin; endostatin; interferon; thrombospondin;
KW   interferon-inducible protein; platelet factor 4; anti-angiogenic;
KW   anti-tumor; multifunctional protein; angiogenic-mediated disease; cancer;
KW   diabetic retinopathy; macular degeneration; arthritis;
KW   tumor cell production.
XX
OS   Homo sapiens.
XX
PN   WO9916889-A1.
XX
PD   08-APR-1999.
XX
PF   30-SEP-1998;   98WO-US020464.
XX
PR   01-OCT-1997;   97US-0060609P.
XX
PA   (SEAR ) SEARLE & CO G D.
XX
PI   Bolanowski MA,  Caparon MH,  Casperson GF,  Gregory SA,  Klein BK;
PI   Mckearn JP;
XX
DR   WPI; 1999-255098/21.
XX
PT   New multifunctional proteins useful for treating angiogenic-mediated
PT   diseases.
XX
PS   Disclosure; Page 106-107; 121pp; English.
XX
CC   The specification describes multifunctional proteins which comprise
CC   combinations of angiostatin, endostatin, interferon, thrombospondin,
CC   interferon-inducible protein and platelet factor 4, and have anti-
CC   angiogenic and/or anti-tumor activity. The multifunctional protein may
CC   exhibit useful properties such as having similar or greater biological
CC   activity when compared to a single factor or by having improved half-life
CC   or decreased adverse side effects, or a combination of these properties.
CC   The proteins can be used for treating an angiogenic-mediated disease,
CC   e.g. cancer, diabetic retinopathy, macular degeneration, or arthritis.
CC   They can also be used for inhibiting the production of tumor cells
CC   (characteristic of lung, breast, ovarian, prostate, pancreatic, gastric,
CC   colon, renal, bladder cancers; melanoma, hepatoma, sarcoma and lymphoma)
CC   in a patient and for inhibiting tumor growth. The present sequence is
CC   used in the course of the invention
XX
SQ   Sequence 183 AA;

  Query Match             100.0%;  Score 893;  DB 2;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 10
AAY08693
ID   AAY08693 standard; protein; 183 AA.
XX
AC   AAY08693;
XX
DT   10-AUG-1999  (first entry)
XX
DE   Human endostatin protein fragment.
XX
KW   Plasminogen; human; angiostatin; endostatin; gene therapy; vector;
KW   anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
KW   tumour growth; solid tumour; diabetic retinopathy; retina.
XX
OS   Homo sapiens.
XX
PN   WO9926480-A1.
XX
PD   03-JUN-1999.
XX
PF   20-NOV-1998;   98WO-US024950.
XX
PR   20-NOV-1997;   97US-00975424.
XX
PA   (GENE-) GENETIX PHARM INC.
PA   (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX
PI   Leboulch P,  Pawliuk RJ,  Bachelot T;
XX
DR   WPI; 1999-357696/30.
DR   N-PSDB; AAX77719.
XX
PT   Anti-angiogenic gene therapy vectors.
XX
PS   Disclosure; Page 74-75; 83pp; English.
XX
CC   This invention describes a novel viral gene therapy vector comprising a
CC   nucleic acid molecule encoding an anti-angiogenic polypeptide chosen from
CC   human or murine angiostatin, human or murine endostatin and angiogenesis-
CC   inhibiting fusions and fragments, where the viral vector is sufficiently
CC   attenuated for use in human gene therapy. The products of the invention
CC   have anti-angiogenic, cytostatic, anti-diabetic and ophthalmological
CC   activity. The vector is used in gene therapy for inhibiting tumour growth
CC   in humans harbouring a solid tumour. The vector expresses an anti-
CC   angiogenic polypeptide. An additional use comprises treatment of diabetic
CC   retinopathy, where the anti-angiogenic polypeptide inhibits angiogenesis
CC   in the vicinity of the retina. The vector is administered to cells ex
CC   vivo and then administered to the patient
XX
SQ   Sequence 183 AA;
```

```
  Query Match             100.0%;  Score 893;  DB 2;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182
```

```
RESULT 11
AAY70252
ID   AAY70252 standard; protein; 183 AA.
XX
AC   AAY70252;
XX
DT   06-JUN-2000  (first entry)
XX
DE   Human angiogenesis inhibitor, endostatin.
XX
KW   Human; immunoglobulin gamma Fc fragment; endostatin; immunofusin;
KW   angiogenesis; inhibitor; cytostatic; antirheumatoid; antiarthritic;
KW   antipsoriatic; antidiabetic; ophthalmological; immunosuppressant;
KW   vasotropic; vulnerary; treatment; antiarteriosclerosis; tumour;
KW   metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;
KW   ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW   myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW   wound granulation; keloid scar; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200011033-A2.
XX
PD   02-MAR-2000.
XX
PF   25-AUG-1999;   99WO-US019329.
XX
PR   25-AUG-1998;   98US-0097883P.
XX
PA   (LEXI-) LEXINGEN PHARM CORP.
XX
PI   Lo K,  Li Y,  Gillies SD;
XX
DR   WPI; 2000-237616/20.
DR   N-PSDB; AAZ51291.
XX
PT   Novel fusion protein of angiostatin or endostatin and an immunoglobulin
PT   FC region, useful for treating conditions mediated by angiogenesis, such
PT   as rheumatoid arthritis, tumors and macular degeneration.
XX
PS   Example 1; Page 41-42; 68pp; English.
XX
CC   The patent discloses a DNA molecule encoding a fusion protein comprising
CC   a signal sequence, an immunoglobulin Fc region, and an angiogenesis
CC   inhibitor selected from angiostatin, endostatin, a plasminogen fragment
CC   having angiostatin activity, a collagen XVIII fragment having endostatin
CC   activity, or combinations of them. The fusion protein (immunofusin) is
CC   used to inhibit angiogenesis and to treat diseases or conditions mediated
CC   by angiogenesis. Conditions that may be treated include solid tumours,
CC   blood born tumours, tumour metastasis, benign tumours including
CC   haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic
CC   granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases
CC   e.g. diabetic retinopathy, retinopathy of prematurity, macular
CC   degeneration, corneal graft rejection, neovascular glaucoma, retrolental
CC   fibroplasia, rubeosis and Osler-Webber syndrome; myocardial angiogenesis,
CC   plaque neovascularisation, telangiectasia, haemophiliac joints'
CC   angiofibroma, wound granulation, and excessive or abnormal stimulation of
CC   endothelial cells, intestinal cells, atherosclerosis, sclerodermal and
CC   hypertrophic scars, i.e. keloid scars. The DNA constructs may be used in
CC   gene therapy. The present sequence is a human endostatin used in the
CC   construction of immunofusin containing human immunoglobulin gamma (IgG)
CC   Fc fragment
XX
SQ   Sequence 183 AA;
```

```
  Query Match             100.0%;  Score 893;  DB 3;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182
```

```
RESULT 12
AAY90771
ID   AAY90771 standard; protein; 183 AA.
XX
AC   AAY90771;
XX
DT   22-AUG-2000  (first entry)
XX
DE   Human angiogenesis inhibiting factor 1 protein.
XX
KW   Human; angiogenesis inhibiting factor 1; IAF-1; tumour; antibody;
KW   abnormal vessel disease.
XX
OS   Homo sapiens.
XX
PN   CN1244536-A.
XX
PD   16-FEB-2000.
XX
PF   10-AUG-1998;   98CN-00117150.
XX
PR   10-AUG-1998;   98CN-00117150.
XX
PA   (ONCO-) INST ONCOLOGY UNDER TUMOR HOSPITAL CHINE.
XX
PI   Yang Z,  Guo W;
XX
DR   WPI; 2000-388168/34.
DR   N-PSDB; AAA29884.
XX
PT   Angiogenesis inhibiting factor 1 and its derivative useful for treating
PT   tumors.
XX
PS   Claim 1; Fig 5; 41pp; Chinese.
XX
CC   The present sequence represents an angiogenesis inhibiting factor (I),
CC   designated IAF-1. The present invention also describes: (1) preparation
CC   of (I) and its derivative; (2) an IAF binding acceptor and its
CC   preparation; and (3) an IAF antibody. (I) is useful for preparing new
CC   biological preparations for effectively treating various tumours and
CC   abnormal-vessel diseases. The IAF antibody is preferably a polyclonal
CC   antibody, mosaic antibody, single stranded antibody and human originated
CC   antibody
XX
SQ   Sequence 183 AA;
```

```
  Query Match             100.0%;  Score 893;  DB 3;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 13
AAB16451
ID   AAB16451 standard; protein; 183 AA.
XX
AC   AAB16451;
XX
DT   27-OCT-2000  (first entry)
XX
DE   Human endostatin protein sequence.
XX
KW   Angiogenesis-inhibiting protein receptor; angiogenesis; angiostatin;
KW   endostatin; plasminogen; laminin; treatment; wound healing; solid tumour;
KW   psoriasis; scleroderma; myocardial angiogenesis; Crohn's disease;
KW   cerebral collateral; arteriovenous malformation; rubeosis; cancer;
KW   diabetic retinopathy; arthritis; wound healing; peptic ulcer;
KW   Helicobacter related disease; fracture; cat scratch fever.
XX
OS   Homo sapiens.
XX
PN   WO200032631-A2.
XX
PD   08-JUN-2000.
XX
PF   06-DEC-1999;   99WO-US028897.
XX
PR   04-DEC-1998;   98US-00206059.
XX
PA   (ENTR-) ENTREMED INC.
XX
PI   Macdonald NJ,  Sim KL;
XX
DR   WPI; 2000-412290/35.
XX
PT   New angiogenesis-inhibiting protein receptors, useful in methods for
PT   treating diseases and processes that are mediated by angiogenesis, such
PT   as solid tumors, psoriasis, scleroderma and myocardial angiogenesis.
XX
PS   Disclosure; Fig 3; 100pp; English.
XX
CC   This invention relates to angiogenesis-inhibiting protein receptors, and
CC   the DNA sequences encoding them. Angiogenesis is the generation of new
CC   blood vessels into a tissue, and normally occurs in wound healing, foetal
CC   and embryonal development and the formation of the corpus luteum,
CC   endometrium and placenta. Angiostatin is a protein (see AAB16450 and
CC   AAA68202) involved in angiogenesis, and has an amino acid sequence
CC   similar to that of a plasminogen fragment (see murine plasminogen
CC   AAB16490). Angiostatin has the ability to inhibit angiogenesis.
CC   Endostatin is also an angiogenesis inhibiting protein (see AAB16451 and
CC   AAA68203). Sequences AAA68242 and AAB16522 represent coding and protein
CC   sequences of human laminin. Laminin is an angiostatin binding protein,
CC   and some of the peptides of the invention share homology with regions of
CC   laminin. Peptides AAB16452-B16521 (excluding AAB16490) are the
CC   angiogenesis-inhibiting protein receptor fragments of the invention. The
CC   peptides bind either angiostatin or endostatin and can be used in methods
CC   for treating diseases and processes that are mediated by angiogenesis,
CC   such as solid tumours, psoriasis, scleroderma, myocardial angiogenesis,
CC   Crohn's disease, cerebral collaterals, arteriovenous malformations,
CC   rubeosis, diabetic retinopathy, arthritis, wound healing, peptic ulcers,
CC   Helicobacter related diseases, fractures, placentation and cat scratch
CC   fever. They are useful for the detection and prognosis of cancer. DNA
CC   sequences A628204-A628241 encode the peptides of the invention
XX
SQ   Sequence 183 AA;

  Query Match             100.0%;  Score 893;  DB 3;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 14
AAB30493
ID   AAB30493 standard; protein; 183 AA.
XX
AC   AAB30493;
XX
DT   06-MAR-2001  (first entry)
XX
DE   Amino acid sequence of human endostatin encoded by plasmid pMALcH#15.
XX
KW   Streptomyces sp. strain C5; SnpA; S. venezuelae; alpha-amylase;
KW   endostatin; cancer; tumour growth; angiogenesis.
XX
OS   Homo sapiens.
XX
PN   WO200060945-A1.
XX
PD   19-OCT-2000.
XX
PF   12-APR-2000; 2000WO-US009747.
XX
PR   13-APR-1999;   99US-0129084P.
XX
PA   (MERI ) MERCK & CO INC.
XX
PI   Desanti CL,  Strohl WR;
XX
DR   WPI; 2000-686970/67.
DR   N-PSDB; AAC62023.
XX
PT   Preparation of soluble recombinant endostatin involves transforming
PT   Streptomycete host with expression vector comprising nucleotide sequence
PT   encoding endostatin operably linked to linker and leader peptide.
XX
PS   Example 1; Fig 6; 57pp; English.
XX
CC   The present sequence represents human endostatin. The protein is
CC   expressed in Streptomyces. Leader sequences of Streptomyces sp. strain C5
CC   SnpA and S. venezuelae alpha-amylase proteins are linked to the N-
CC   terminal of endostatin. This ensures that endostatin protein is produced
CC   as a secreted, soluble protein which needs no refolding, is stable in the
CC   fermentation broth and is produced in large quantities. The method is
CC   used for preparing soluble recombinant human, murine or primate
CC   endostatin, which is useful in the treatment of cancer, inhibition of
CC   tumour growth, inhibition of angiogenesis, isolation of receptors for
CC   endostatin and for identification of anti-angiogenic compounds in assays.
CC   The endostatin protein is produced as a secreted, soluble protein which
CC   needs no refolding, is stable in the fermentation broth and is produced
```

```
CC   in large quantities. Streptomycetes are amenable for cultivation in large
CC   fermentations allowing for large quantities of soluble endostatin to be
CC   produced
XX
SQ   Sequence 183 AA;

  Query Match             100.0%;  Score 893;  DB 3;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 15
AAB49379
ID   AAB49379 standard; protein; 183 AA.
XX
AC   AAB49379;
XX
DT   02-MAR-2001  (first entry)
XX
DE   Human endostatin SEQ ID NO: 2.
XX
KW   Endostatin; antiangiogenic; angiogenesis; human; mouse; chicken; cancer;
KW   inflammation; angiogenesis-dependent disease.
XX
OS   Homo sapiens.
XX
PN   WO200067771-A1.
XX
PD   16-NOV-2000.
XX
PF   02-MAY-2000; 2000WO-US012063.
XX
PR   06-MAY-1999;   99US-0132907P.
PR   14-JUL-1999;   99US-00353333.
XX
PA   (BURN-) BURNHAM INST.
XX
PI   Vuori K;
XX
DR   WPI; 2001-040937/05.
DR   N-PSDB; AAC88289.
XX
PT   Endostatin peptide comprising at least four endostatin amino acid
PT   residues are e.g. angiogenesis inhibitors for treating cancer and
PT   diabetic retinopathy.
XX
PS   Disclosure; Fig 1; 146pp; English.
XX
CC   The present invention provides endostatin peptides which can be used in
CC   the modulation of angiogenesis. This is useful in the treatment of
CC   cancers, inflammation, rheumatoid arthritis, chronic articular
CC   rheumatism, psoriasis, disorders associated with inopportune invasion of
CC   vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy
CC   of prematurity, macular degeneration, corneal graft rejection,
CC   retrolental fibroplasia, rubeosis, capillary proliferation in
CC   atherosclerotic plaques and osteoporosis. Other angiogenesis-dependent
CC   diseases include Osler-Webber syndrome, myocardial angiogenesis, plaque
CC   neovasculisation, telangiectasia, haemophiliac joints and wound
CC   granulation. In addition, the peptides can be used as birth control
CC   agents
XX
SQ   Sequence 183 AA;

  Query Match             100.0%;  Score 893;  DB 4;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 16
AAU00896
ID   AAU00896 standard; protein; 183 AA.
XX
AC   AAU00896;
XX
DT   04-JUL-2001  (first entry)
XX
DE   Human Endostatin(TM) protein.
XX
KW   Human; Endostatin(TM); angiogenesis mediated disease; solid tumours;
KW   blood borne tumour; leukaemia; tumour metastasis; benign tumour;
KW   haemangioma; acoustic neuroma; neurofibroma; trachoma; rubeosis;
KW   pyogenic granuloma; rheumatoid arthritis; psoriasis; colon cancer;
KW   ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW   retinopathy of prematurity; macular corneal graft rejection;
KW   neovascular glaucoma; retrolental fibroplasia; Osler-Webber Syndrome;
KW   myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW   haemophiliac joint; angiofibroma; wound granulation.
XX
OS   Homo sapiens.
XX
PN   WO200119989-A2.
XX
PD   22-MAR-2001.
XX
PF   14-SEP-2000; 2000WO-US025166.
XX
PR   14-SEP-1999;   99US-0153698P.
XX
PA   (ENTR-) ENTREMED INC.
XX
PI   Liang H,  Sim KL,  Chang-Murad A,  Zhou X,  Madsen J,  Boerner RJ;
PI   Bermejo LL,  Mistry FR,  Shepard SR,  Schrimsher JL;
XX
DR   WPI; 2001-244802/25.
DR   N-PSDB; AAS00867.
XX
PT   Producing Endostatin protein for treating angiogenesis mediated diseases
PT   such as solid tumors, comprises recombinantly producing the protein using
PT   an expression system, and recovering and purifying the protein.
XX
PS   Claim 5; Page 29; 67pp; English.
XX
CC   The sequence represents Human Endostatin(TM). The new method of the
CC   invention is useful for producing, recovering and purifying Endostatin
CC   (TM) from biological sources, such as biological fluids, tissues, cells,
CC   culture media, and fermentation media. Endostatin(TM) is useful for
CC   treating angiogenesis mediated diseases such as solid tumours, blood
CC   borne tumours, leukaemias, tumour metastases, benign tumours, e.g.
CC   haemangioma, acoustic neuromas, neurofibromas, trachomas, and pyogenic
CC   granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases,
CC   e.g., diabetic retinopathy, retinopathy of prematurity, macular
CC   degeneration, corneal graft rejection, neovascular glaucoma, colon
```

```
CC   cancer, retrolental fibroplasia, rubeosis, Osler-Webber Syndrome,
CC   myocardial angiogenesis, plaque neovascularisation, telangiectasia,
CC   haemophiliac joints, angiofibroma, and wound granulation. Endostatin(TM)
CC   is also useful for treating disease of excessive or abnormal stimulation
CC   of endothelial cells such as intestinal adhesions, atherosclerosis,
CC   scleroderma and hypertrophic scars. Higher yields of more purified, and
CC   biologically active Endostatin(TM) are obtained by the new method.
CC   Endostatin(TM) can be stored in buffers for extended periods of time, and
CC   also subjected to lyophilisation, while preserving biological activity.
CC   Centrifugation of broth from fermentation steps in production is avoided,
CC   preventing unwanted potential cellular lysis and contamination with
CC   additional proteins, pigments, enzymes and other cellular chemicals and
CC   debris
XX
SQ   Sequence 183 AA;

  Query Match             100.0%;  Score 893;  DB 4;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 17
ABB79901
ID   ABB79901 standard; protein; 183 AA.
XX
AC   ABB79901;
XX
DT   05-DEC-2002  (first entry)
XX
DE   Human endostatin polypeptide.
XX
KW   Endostatin; human; ophthalmological; ocular neovascularisation;
KW   choroidal neovascularisation; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200267971-A2.
XX
PD   06-SEP-2002.
XX
PF   21-FEB-2002; 2002WO-US005336.
XX
PR   22-FEB-2001; 2001US-0270787P.
PR   04-APR-2001; 2001US-0281296P.
XX
PA   (NOVS ) NOVARTIS AG.
XX
PI   Brazzell RK,  Campochiaro PA,  Dixon KH;
XX
DR   WPI; 2002-698636/75.
DR   N-PSDB; ABQ81193.
XX
PT   Treating or preventing choroidal neovascularization comprises increasing
PT   the amount of endostatin in ocular tissues of afflicted individuals to a
PT   choroidal neovascularization inhibiting level.
XX
PS   Claim 2; Page 39; 44pp; English.
XX
CC   The present sequence is the protein sequence of a human endostatin
CC   polypeptide. A claimed method for the treatment of ocular
CC   neovascularisation, especially choroidal neovascularisation, involves
CC   increasing the level of endostatin in ocular tissue, especially where the
CC   endostatin comprises the present sequence, or is its fragment, derivative
CC   or variant. The increase is effected by administering a viral vector,
CC   especially an adenovirus, adeno-associated virus, a retrovirus or
CC   lentivirus vector, comprising an endostatin-encoding nucleic acid. Cells
CC   secreting endostatin may be encapsulated and implanted within an
CC   individual. The method is used when ocular neovascularisation is caused
CC   by histoplasmosis, pathological myopia, angioid streaks, anterior
CC   ischaemic optic neuropathy, bacterial endocarditis, Best's disease,
CC   birdshot retinochoroidopathy, choroidal haemangioma, choroidal naevi,
CC   choroidal nonperfusion, choroidal osteomas, choroidal rupture,
CC   choroideraemia, chronic retinal detachment, coloboma of the retina,
CC   Drusen, endogenous Candida endophthalmitis, extrapapillary hamartoma of
CC   the retinal pigmented epithelium, fundus flavimaculatus, idiopathic,
CC   macular hole, malignant melanoma, membranproliferative glomerulonephritis
CC   (type II), metallic intraocular foreign body, morning glory disc
CC   syndrome, multiple evanescent white-dot syndrome, neovascularisation of
CC   ora serrata, operating microscope burn, optic nerve head pits,
CC   photocoagulation, punctate inner choroidopathy, rubella, sarcoidosis,
CC   serpiginous or geographic choroiditis, subretinal fluid drainage, tiled
CC   disc syndrome, Toxoplasma retinochoroiditis, tuberculosis, Vogt-Koyanagi-
CC   Harada syndrome, diabetic retinopathy, non-diabetic retinopathy, brain
CC   vein occlusion, central retinal vein occlusion, retinopathy in premature
CC   infants, rubeosis iridis, neovascular glaucoma, perifoveal
CC   telangiectasis, sickle cell retinopathy, Eale's disease, retinal
CC   vasculitis, Von Hippel Lindau disease, radiation retinopathy, retinal
CC   cryoinjury, retinitis pigmentosa, retinochoroidal coloboma, corneal
CC   neovascularisation due to herpes simplex keratitis, corneal ulcers,
CC   keratoplasty, pterigyia and trauma (all claimed)
XX
SQ   Sequence 183 AA;

  Query Match             100.0%;  Score 893;  DB 5;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 18
AAM49503
ID   AAM49503 standard; protein; 183 AA.
XX
AC   AAM49503;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Human endostatin protein.
XX
KW   Endostatin; human; proliferation; blood vessel endothelium; regeneration;
KW   tumour; blood vessel; treatment; amplification.
XX
OS   Homo sapiens.
XX
PN   CN1177005-A.
XX
PD   25-MAR-1998.
XX
PF   10-SEP-1997;   97CN-00107112.
XX
PR   10-SEP-1997;   97CN-00107112.
XX
PA   (XUGG/) XU G.
```

XX
PI Xu G, Ren M, Xu L;
XX
DR WPI; 2002-106746/15.
DR N-PSDB; ABA99261.
XX
PT Gene clone of inhibitory factor for hyperplasia of inner blood vessel
PT cells in human body's real tumor, and its use in anti-tumor blood vessel
PT regeneration.
XX
PS Disclosure; Page 4 (Disclosure); 6pp; Chinese.
XX
CC This invention describes a novel preparation which inhibits the
CC proliferation of blood vessel endothelium and prevents the regeneration
CC activity of tumour blood vessels. The preparation can also be used as a
CC biological preparation in the treatment of tumours. This sequence
CC represents the human endostatin protein described in the invention
XX
SQ Sequence 183 AA;

Query Match 100.0%; Score 893; DB 5; Length 183;
Best Local Similarity 100.0%; Pred. No. 3.2e-101;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
Db 13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy 61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
Db 73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy 121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
Db 133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182

RESULT 19
AAM48895
ID AAM48895 standard; protein; 183 AA.
XX
AC AAM48895;
XX
DT 04-APR-2002 (first entry)
XX
DE Human endostatin protein.
XX
KW Human; angiostatin; endostatin; angiogenesis; cancer; metastasis;
KW psoriasis; scleroderma; Crohn's disease; corneal disease; retinopathy;
KW arthritis; wound healing; Helicobacter pylori; peptic ulcer;
KW gene therapy; angiostatin antagonist; endostatin antagonist;
KW antiangiogenic; cytostatic; antiarthritic; antiinflammatory;
KW cerebroprotective; antidiabetic; virucide; antipyretic; vulnerary;
KW gynaecological; cat scratch fever.
XX
OS Homo sapiens.
XX
PN WO200193897-A2.
XX
PD 13-DEC-2001.
XX
PF 04-JUN-2001; 2001WO-US017947.
XX
PR 02-JUN-2000; 2000US-0209065P.
PR 08-MAY-2001; 2001US-0289387P.
XX
PA (ENTR-) ENTREMED INC.
XX
PI Sim KL, Macdonald NJ;
XX
DR WPI; 2002-130569/17.
XX
PT Regulating angiogenesis and treatment of angiogenesis-mediated diseases,
PT e.g. hemangioma, tumors or cancer, by administering a tropomyosin binding
PT compound or actin disrupting compound.
XX
PS Disclosure; Fig 3; 95pp; English.
XX
CC The present invention relates to methods of regulating angiogenesis in an
CC individual by administering an angiogenesis regulating composition
CC comprising a tropomyosin binding compound or an actin disrupting
CC compound. The compositions are useful for treating diseases and processes
CC mediated by angiogenesis including haemangioma, solid tumours, blood
CC bourne tumours, leukaemia, metastasis, Crohn's disease, coronary or
CC cerebral collaterals, arthritis, diabetic neovascularisation, macular
CC degeneration, wound healing, Helicobacter related diseases, ovulation,
CC menstruation, and cat scratch fever. The present sequence is a protein
CC described in the exemplification of the invention
XX
SQ Sequence 183 AA;

Query Match 100.0%; Score 893; DB 5; Length 183;
Best Local Similarity 100.0%; Pred. No. 3.2e-101;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
Db 13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy 61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
Db 73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy 121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
Db 133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182

RESULT 20
AAU97132
ID AAU97132 standard; protein; 183 AA.
XX
AC AAU97132;
XX
DT 13-AUG-2002 (first entry)
XX
DE Human endostatin.
XX
KW Human; angiogenesis; PITSLRE protein kinase; cancer; arthritis;
KW macular degeneration; diabetic retinopathy; angiogenic-related disease;
KW haemangioma; blood borne tumour; leukaemia; neovascularisation;
KW coronary collateral; cerebral collateral; neovascular glaucoma;
KW corneal disease; wound healing; Helicobacter related disease; fracture;
KW keloid; ovulation; menstruation.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 64. .126
FT /note= "Internal peptide of endostatin is a PITSLRE
FT homologous region"
XX
PN WO200230982-A2.
XX
PD 18-APR-2002.
XX
PF 15-OCT-2001; 2001WO-US032437.
XX
PR 13-OCT-2000; 2000US-0240127P.
XX
PA (ENTR-) ENTREMED INC.
XX
PI Sim KL, Liang H;
XX
DR WPI; 2002-435440/46.
DR N-PSDB; ABK50685.

```
XX
PT   Regulating angiogenesis for treating scleroderma, leukemia, keloids by
PT   administering a protein that is homologous to PITSLRE protein kinase and
PT   an angiogenic factor or a protein kinase and its active fragments.
XX
PS   Disclosure; Fig 2A; 45pp; English.
XX
CC   The present invention relates to methods and compositions of inhibiting
CC   angiogenesis. The method comprises administering to a human or animal a
CC   composition comprising a protein that is homologous to PITSLRE protein
CC   kinases (PK) and an angiogenic factor. The method is useful for
CC   regulating angiogenesis related to cancer, arthritis, macular
CC   degeneration, and diabetic retinopathy. The compositions are useful for
CC   inhibiting angiogenic-related diseases. The method and compositions are
CC   useful in treating diseases and processes that are mediated by
CC   angiogenesis including haemangioma, solid tumours, blood borne tumours,
CC   leukaemia, metastasis, telangiectasia, psoriasis, scleroderma, pyogenic
CC   granuloma, myocardial angiogenesis, Crohn's disease, plaque
CC   neovascularisation, coronary collaterals, cerebral collaterals,
CC   arteriovenous malformations, ischaemic limb angiogenesis, corneal
CC   diseases, rubeosis, neovascular glaucoma, diabetic retinopathy,
CC   retrolental fibroplasia, arthritis, diabetic neovascularisation, macular
CC   degeneration, wound healing, peptic ulcer, Helicobacter related diseases,
CC   fractures, keloids, vasculogenesis, haematopoiesis, ovulation,
CC   menstruation, placentation, and cat scratch fever. The method of the
CC   invention provides a therapy for cancer that has minimal side effects.
CC   The present sequence represents human endostatin which is used to
CC   generate angiogenesis-inhibiting peptides
XX
SQ   Sequence 183 AA;

  Query Match             100.0%;  Score 893;  DB 5;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 21
AAG79753
ID   AAG79753 standard; protein; 183 AA.
XX
AC   AAG79753;
XX
DT   18-MAR-2003  (first entry)
XX
DE   Human endostatin.
XX
KW   Human; plasminogen; angiostatin; neovascularisation; kringle domain;
KW   cell proliferation; viral vector; replication-defective; cancer; tumour.
XX
OS   Homo sapiens.
XX
PN   WO200288173-A2.
XX
PD   07-NOV-2002.
XX
PF   29-APR-2002; 2002WO-US013461.
XX
PR   30-APR-2001; 2001US-0287673P.
PR   05-APR-2002; 2002US-0370634P.
XX
PA   (CELL-) CELL GENESYS INC.
XX
PI   Chang B,  Wu WW,  Macarthur J,  Patel S,  Jooss K,  Mendez M;
XX
DR   WPI; 2003-129131/12.
DR   N-PSDB; ABA00774.
XX
PT   New recombinant viral vector expressing human angiostatin useful for
PT   inhibiting angiogenesis in a mammalian subject with cancer or tumor.
XX
PS   Example 4; Page 82-83; 83pp; English.
XX
CC   This sequence represents endostatin. Endostatin is a 20 kD C-terminal
CC   fragment of collagen XVIII that inhibits angiogenesis. The endostatin
CC   coding sequence may be used in the recombinant viral vector of the
CC   invention for obtaining angiostatin activity. The vector comprises a
CC   promoter capable of expressing human angiostatin operably linked to a
CC   structural gene encoding one or more domains of human angiostatin. The
CC   vector, which may be a replication-defective viral vector, is useful for
CC   inhibiting angiogenesis in a mammal, especially with cancer or a tumour
XX
SQ   Sequence 183 AA;

  Query Match             100.0%;  Score 893;  DB 6;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 22
AAU76690
ID   AAU76690 standard; protein; 193 AA.
XX
AC   AAU76690;
XX
DT   21-MAY-2002  (first entry)
XX
DE   Synthetic plasmid pEnd-HR#2 FPD fusion protein sequence.
XX
KW   Mouse; Ig signal peptide; mIgSP; functional protein domain; FPD;
KW   primary translational product; PTP; DNA construct; regulatory DNA;
KW   DNA targeting segment; regulatory factor; single regulatory unit;
KW   monoclonal antibody; recombination-derived alteration; blood product;
KW   human; COL18A1; mutant; mutein; fusion protein.
XX
OS   Mus sp.
OS   Homo sapiens.
OS   Synthetic.
OS   Chimeric.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .19
FT                   /label= Signal_peptide
FT                   /note= "Mouse Ig signal peptide (mIgSP)"
FT   Region          1. .19
FT                   /note= "Encoded by mouse Ig signal peptide (mIgSP) exon"
FT   Protein         20. .193
FT                   /label= Mature_human_COL18A1_protein"
FT                   /note= "Contains exons 39-41"
FT   Region          20. .84
FT                   /note= "Encoded by human COL18A1 exon 39"
FT   Region          85. .124
FT                   /note= "Encoded by human COL18A1 exon 40"
```

```
FT   Region          125. .193
FT                   /note= "Encoded by human COL18A1 exon 41"
XX
PN   WO200210372-A1.
XX
PD   07-FEB-2002.
XX
PF   01-AUG-2001; 2001WO-GB003455.
XX
PR   01-AUG-2000; 2000GB-00018876.
XX
PA   (ISTF ) ARS APPLIED RES SYSTEMS HOLDING NV.
PA   (CHAP/) CHAPMAN P W.
XX
PI   Chapman PW,  De Luca G,  Falciola L;
XX
DR   WPI; 2002-195963/25.
DR   N-PSDB; ABK09978.
XX
PT   Producing functional protein domain by growing host cell transfected with
PT   DNA construct having regulatory DNA and DNA targeting segment, and
PT   optionally culturing homologously recombinant cell and collecting
PT   protein.
XX
PS   Example; Fig 9; 116pp; English.
XX
CC   The present invention relates to a new method of producing a protein,
CC   such as functional protein domain, that is either C- or N-terminus of the
CC   primary translational product (PTP) of a gene, where the protein has
CC   biological activity which is distinct from PTP. The method of the
CC   invention involves growing a host cell transfected with a DNA construct
CC   comprising a regulatory DNA and a DNA targeting segment. This method is
CC   useful for producing a functional protein domain of proteins such as
CC   regulatory factors, blood products and monoclonal antibodies. The method
CC   described in the invention allows controlled and precise modification of
CC   the host cell genome in order to produce functional protein domain (FPD).
CC   The amount of exogenous sequence to be integrated in the host cell genome
CC   is very limited since, as coding sequence, the original coding sequence
CC   present in the host cell genome itself is used. Use of the host cell
CC   sequence encoding FPD also provides the advantages of both eliminating
CC   any recombination-derived alteration of such coding sequence, and also
CC   making use of the same post-transcriptional (e.g., splicing) and/or post-
CC   translational (e.g. glycosylation, phosphorylation) processes that are
CC   actually applied in vivo for the maturation of FPD. The use of a single
CC   regulatory unit eliminates the necessity of manipulating the
CC   complementary DNA coding for the PTP to isolate the segment coding for
CC   the FPD, and adapt it to the expression vector. The present amino acid
CC   sequence represents the plasmid pEnd-HR#2 FPD fusion protein of the
CC   invention. This fusion protein contains the mouse Ig signal peptide
CC   (mIgSP) sequence fused to exons 39-41 of the human COL18A1 sequence
XX
SQ   Sequence 193 AA;

  Query Match             100.0%;  Score 893;  DB 5;  Length 193;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 82

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 142

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 192


RESULT 23
AAW90874
ID   AAW90874 standard; protein; 195 AA.
XX
AC   AAW90874;
XX
DT   07-JUL-2000  (first entry)
XX
DE   Human HMW endostatin (1) protein.
XX
KW   Endostatin; human; renal insufficiency; antitumor; antiproliferative;
KW   treatment; angiogenesis; tumor; vascular disease.
XX
OS   Homo sapiens.
XX
PN   WO200017240-A1.
XX
PD   30-MAR-2000.
XX
PF   21-SEP-1999;   99WO-EP006963.
XX
PR   21-SEP-1998;   98DE-01042992.
PR   03-APR-1999;   99DE-01015267.
PR   08-JUN-1999;   99DE-01026040.
XX
PA   (HAEM-) HAEMOPEP PHARMA GMBH.
XX
PI   Staendker L,  Forssmann W;
XX
DR   WPI; 2000-292826/25.
XX
PT   New high molecular weight form of endostatin, useful e.g. as
PT   antiangiogenic agent for treating cancer, isolated from hemofiltrate of
PT   patients with kidney failure.
XX
PS   Claim 2; Page 18-19; 32pp; German.
XX
CC   This invention describes novel human high molecular weight (HMW)
CC   endostatin (hE) protein fragments isolated from the hemofiltrate of
CC   patients with renal insufficiency. The products of the invention have
CC   antitumor and antiproliferative activity. hE is used to treat; (i)
CC   diseases that involve uncontrolled angiogenesis, particularly tumors; and
CC   (ii) vascular diseases of supporting or connective tissue, respiratory
CC   tract, cardiovascular system, urogenital tract and nervous system, or
CC   sensory organs (particularly the eye). hE is also used to raise specific
CC   antibodies which are used for diagnosis and treatment of conditions that
CC   involve overexpression of hE. hE has a very long plasma half-life and can
CC   be administered repeatedly without inducing an immune response. AAW90874-
CC   W90877 represent the endostatin proteins described in the method of the
CC   invention
XX
SQ   Sequence 195 AA;

  Query Match             100.0%;  Score 893;  DB 3;  Length 195;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         26 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 85

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 145

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        146 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 195


RESULT 24
AAB30495
ID   AAB30495 standard; protein; 216 AA.
XX
AC   AAB30495;
XX
```

```
DT   06-MAR-2001  (first entry)
XX
DE   Amino acid sequence of vaa-endostatin fusion protein in pANT3052.
XX
KW   Streptomyces sp. strain C5; SnpA; S. venezuelae; alpha-amylase;
KW   endostatin; cancer; tumour growth; angiogenesis.
XX
OS   Synthetic.
OS   Streptomyces sp.
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .28
FT                   /note= "vaa signal sequence"
FT   Protein         29. .216
FT                   /note= "endostatin"
XX
PN   WO200060945-A1.
XX
PD   19-OCT-2000.
XX
PF   12-APR-2000; 2000WO-US009747.
XX
PR   13-APR-1999;   99US-0129084P.
XX
PA   (MERI ) MERCK & CO INC.
XX
PI   Desanti CL,  Strohl WR;
XX
DR   WPI; 2000-686970/67.
DR   N-PSDB; AAC62025.
XX
PT   Preparation of soluble recombinant endostatin involves transforming
PT   Streptomycete host with expression vector comprising nucleotide sequence
PT   encoding endostatin operably linked to linker and leader peptide.
XX
PS   Example 1; Fig 10A-B; 57pp; English.
XX
CC   The present sequence represents a fusion protein of vaa and endostatin.
CC   The specification describes a method for the production of soluble,
CC   recombinant human endostatin in Streptomyces. Leader sequences of
CC   Streptomyces sp. strain C5 SnpA and S. venezuelae alpha-amylase proteins
CC   are linked to the N-terminal of endostatin. This ensures that endostatin
CC   protein is produced as a secreted, soluble protein which needs no
CC   refolding, is stable in the fermentation broth and is produced in large
CC   quantities. The method is used for preparing soluble recombinant human,
CC   murine or primate endostatin, which is useful in the treatment of cancer,
CC   inhibition of tumour growth, inhibition of angiogenesis, isolation of
CC   receptors for endostatin and for identification of anti-angiogenic
CC   compounds in assays. The endostatin protein is produced as a secreted,
CC   soluble protein which needs no refolding, is stable in the fermentation
CC   broth and is produced in large quantities. Streptomycetes are amenable
CC   for cultivation in large fermentations allowing for large quantities of
CC   soluble endostatin to be produced
XX
SQ   Sequence 216 AA;

  Query Match             100.0%;  Score 893;  DB 3;  Length 216;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         46 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 105

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 165

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        166 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 215
```

```
RESULT 25
AAU76689
ID   AAU76689 standard; protein; 275 AA.
XX
AC   AAU76689;
XX
DT   21-MAY-2002  (first entry)
XX
DE   Synthetic plasmid pEnd-HR#1 FPD fusion protein sequence.
XX
KW   Mouse; Ig signal peptide; mIgSP; functional protein domain; FPD;
KW   primary translational product; PTP; DNA construct; regulatory DNA;
KW   DNA targeting segment; regulatory factor; single regulatory unit;
KW   monoclonal antibody; recombination-derived alteration; blood product;
KW   human; COL18A1; mutant; mutein; fusion protein.
XX
OS   Mus sp.
OS   Homo sapiens.
OS   Synthetic.
OS   Chimeric.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .19
FT                   /label= Signal_peptide
FT                   /note= "Mouse Ig signal peptide (mIgSP)"
FT   Region          1. .19
FT                   /note= "Encoded by mouse Ig signal peptide (mIgSP) exon"
FT   Protein         20. .275
FT                   /label= Mature_human_COL18A1_protein"
FT                   /note= "Contains exons 38-41"
XX
PN   WO200210372-A1.
XX
PD   07-FEB-2002.
XX
PF   01-AUG-2001; 2001WO-GB003455.
XX
PR   01-AUG-2000; 2000GB-00018876.
XX
PA   (ISTF ) ARS APPLIED RES SYSTEMS HOLDING NV.
PA   (CHAP/) CHAPMAN P W.
XX
PI   Chapman PW,  De Luca G,  Falciola L;
XX
DR   WPI; 2002-195963/25.
DR   N-PSDB; ABK09977.
XX
PT   Producing functional protein domain by growing host cell transfected with
PT   DNA construct having regulatory DNA and DNA targeting segment, and
PT   optionally culturing homologously recombinant cell and collecting
PT   protein.
XX
PS   Example; Fig 8; 116pp; English.
XX
CC   The present invention relates to a new method of producing a protein,
CC   such as functional protein domain, that is either C- or N-terminus of the
CC   primary translational product (PTP) of a gene, where the protein has
CC   biological activity which is distinct from PTP. The method of the
CC   invention involves growing a host cell transfected with a DNA construct
CC   comprising a regulatory DNA and a DNA targeting segment. This method is
CC   useful for producing a functional protein domain of proteins such as
CC   regulatory factors, blood products and monoclonal antibodies. The method
CC   described in the invention allows controlled and precise modification of
CC   the host cell genome in order to produce functional protein domain (FPD).
CC   The amount of exogenous sequence to be integrated in the host cell genome
CC   is very limited since, as coding sequence, the original coding sequence
CC   present in the host cell genome itself is used. Use of the host cell
CC   sequence encoding FPD also provides the advantages of both eliminating
CC   any recombination-derived alteration of such coding sequence, and also
CC   making use of the same post-transcriptional (e.g., splicing) and/or post-
CC   translational (e.g. glycosylation, phosphorylation) processes that are
CC   actually applied in vivo for the maturation of FPD. The use of a single
```

```
CC   regulatory unit eliminates the necessity of manipulating the
CC   complementary DNA coding for the PTP to isolate the segment coding for
CC   the FPD, and adapt it to the expression vector. The present amino acid
CC   sequence represents the plasmid pEnd-HR#1 FPD fusion protein of the
CC   invention. This fusion protein contains the mouse Ig signal peptide
CC   (mIgSP) sequence fused to exons 38-41 of the human COL18A1 sequence
XX
SQ   Sequence 275 AA;

  Query Match             100.0%;  Score 893;  DB 5;  Length 275;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        105 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 164

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        165 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 224

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        225 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 274


RESULT 26
AAU76688
ID   AAU76688 standard; protein; 310 AA.
XX
AC   AAU76688;
XX
DT   21-MAY-2002  (first entry)
XX
DE   Human collagen XVIII 1alpha NC1 domain protein sequence.
XX
KW   Human; collagen XVIII 1alpha NC1 domain; functional protein domain; FPD;
KW   primary translational product; PTP; DNA construct; regulatory DNA;
KW   DNA targeting segment; regulatory factor; single regulatory unit;
KW   monoclonal antibody; recombination-derived alteration; blood product.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          1. .54
FT                   /label= Multimerisation_domain
FT   Region          1. .43
FT                   /note= "Encoded by exon 36"
FT   Region          44. .54
FT                   /note= "Encoded by exon 37"
FT   Region          55. .136
FT                   /label= Hinge_region
FT                   /note= "Encoded by exon 38"
FT   Domain          137. .310
FT                   /label= Endostatin_core_domain
FT                   /note= "Autonomous folding unit"
FT   Region          137. .201
FT                   /note= "Encoded by exon 39"
FT   Region          202. .241
FT                   /note= "Encoded by exon 40"
FT   Region          242. .310
FT                   /note= "Encoded by exon 41"
XX
PN   WO200210372-A1.
XX
PD   07-FEB-2002.
XX
PF   01-AUG-2001; 2001WO-GB003455.
XX
PR   01-AUG-2000; 2000GB-00018876.
XX
PA   (ISTF ) ARS APPLIED RES SYSTEMS HOLDING NV.
PA   (CHAP/) CHAPMAN P W.
XX
PI   Chapman PW,  De Luca G,  Falciola L;
XX
DR   WPI; 2002-195963/25.
XX
PT   Producing functional protein domain by growing host cell transfected with
PT   DNA construct having regulatory DNA and DNA targeting segment, and
PT   optionally culturing homologously recombinant cell and collecting
PT   protein.
XX
PS   Example; Fig 3; 116pp; English.
XX
CC   The present invention relates to a new method of producing a protein,
CC   such as functional protein domain, that is either C- or N-terminus of the
CC   primary translational product (PTP) of a gene, where the protein has
CC   biological activity which is distinct from PTP. The method of the
CC   invention involves growing a host cell transfected with a DNA construct
CC   comprising a regulatory DNA and a DNA targeting segment. This method is
CC   useful for producing a functional protein domain of proteins such as
CC   regulatory factors, blood products and monoclonal antibodies. The method
CC   described in the invention allows controlled and precise modification of
CC   the host cell genome in order to produce functional protein domain (FPD).
CC   The amount of exogenous sequence to be integrated in the host cell genome
CC   is very limited since, as coding sequence, the original coding sequence
CC   present in the host cell genome itself is used. Use of the host cell
CC   sequence encoding FPD also provides the advantages of both eliminating
CC   any recombination-derived alteration of such coding sequence, and also
CC   making use of the same post-transcriptional (e.g., splicing) and/or post-
CC   translational (e.g. glycosylation, phosphorylation) processes that are
CC   actually applied in vivo for the maturation of FPD. The use of a single
CC   regulatory unit eliminates the necessity of manipulating the
CC   complementary DNA coding for the PTP to isolate the segment coding for
CC   the FPD, and adapt it to the expression vector. The present amino acid
CC   sequence represents the human collagen XVIII 1alpha NC1, a functional
CC   protein domain used to illustrate the method of the invention
XX
SQ   Sequence 310 AA;

  Query Match             100.0%;  Score 893;  DB 5;  Length 310;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 199

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 259

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 309


RESULT 27
ABG73586
ID   ABG73586 standard; protein; 513 AA.
XX
AC   ABG73586;
XX
DT   03-MAR-2003  (first entry)
XX
DE   Human Endostatin/IgG1Fc fusion construct.
XX
KW   Human; endostatin; IgG1Fc; tumour; vascular endothelial proliferation;
KW   vascular endothelial cytopoiesis inhibiting factor; inhibitor;
KW   fusion construct.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
```

FT Misc-difference 20
FT /note= "Encoded by CGT"
FT Misc-difference 271. .277
FT /note= "Encoded by TAGTAA"
XX
PN CN1354186-A.
XX
PD 19-JUN-2002.
XX
PF 30-NOV-2000; 2000CN-00123347.
XX
PR 30-NOV-2000; 2000CN-00123347.
XX
PA (LIAO-) LIAONING WEIXING BIOLOGICAL PROD INST CO.
XX
PI Chen L, Li Z, Liu Q;
XX
DR WPI; 2002-751441/82.
DR N-PSDB; ABQ76740.
XX
PT Preparation of recombinant human vascular endothelial cytopoiesis
PT suppressor factor with human LgG1Fc fragment molecular structure and
PT application of its product.
XX
PS Disclosure; Page 7-8 (Disclosure); 12pp; Chinese.
XX
CC This invention describes a novel method for the preparation of
CC recombinant human vascular endothelial cytopoiesis inhibiting factor with
CC human IgG1Fc fragment molecular structure and its product application.
CC The novel factor is derived from endostatin (using PCR to screen a human
CC foetal kidney cell cDNA library) and human IgG1Fc. The product of the
CC invention can specifically inhibit tumour vascular endothelial
CC proliferation and can be used for curing several tumour types. This
CC sequence represents a fusion construct composed of human endostatin and
CC human IgG1Fc, described in the disclosure of the invention
XX
SQ Sequence 513 AA;

Query Match 100.0%; Score 893; DB 5; Length 513;
Best Local Similarity 100.0%; Pred. No. 1.5e-100;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
Db 100 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 159

Qy 61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
Db 160 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 219

Qy 121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
Db 220 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 269

RESULT 28
ABP41878
ID ABP41878 standard; protein; 682 AA.
XX
AC ABP41878;
XX
DT 22-AUG-2002 (first entry)
XX
DE Human ovarian antigen HEEBK29, SEQ ID NO:3010.
XX
KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antiinflammatory; gynaecological; reproductive; chromosome 21q22.3.
XX
OS Homo sapiens.
XX
PN WO200200677-A1.
XX
PD 03-JAN-2002.
XX
PF 07-JUN-2001; 2001WO-US018569.
XX
PR 07-JUN-2000; 2000US-0209467P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
DR WPI; 2002-147878/19.
DR N-PSDB; ABQ54955.
XX
PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian
PT cancer), immune disorders, cardiovascular disorders and neurological
PT diseases.
XX
PS Claim 11; SEQ ID NO 3010; 2922pp; English.
XX
CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and
CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vaginitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents a human ovarian antigen of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 682 AA;

Query Match 100.0%; Score 893; DB 5; Length 682;
Best Local Similarity 100.0%; Pred. No. 2.2e-100;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
Db 512 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 571

Qy 61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
Db 572 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 631

Qy 121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170

```
Db        632 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 681

RESULT 29
AAW26327
ID   AAW26327 standard; protein; 684 AA.
XX
AC   AAW26327;
XX
DT   19-NOV-1997  (first entry)
XX
DE   Human alpha-1 collagen (XVIII).
XX
KW   Alpha-1 collagen; type XVIII collagen; cartilage degeneration.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .6
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         7. .12
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         13. .18
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         19. .24
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         25. .30
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         31. .36
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         37. .42
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         48. .53
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         54. .59
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         74. .79
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         80. .85
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         86. .91
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         92. .97
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         98. .103
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         104. .109
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         110. .115
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         121. .126
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         129. .134
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         135. .140
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         141. .146
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         147. .152
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         153. .158
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         159. .164
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         165. .170
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         171. .176
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         181. .186
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         187. .192
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         193. .198
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         215. .220
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         221. .226
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         227. .232
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         233. .238
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         239. .244
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         257. .262
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         263. .268
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         269. .274
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         275. .280
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         286. .291
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         292. .297
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         298. .303
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         309. .314
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         315. .320
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         322. .328
FT                   /label= GXYGX'Y'_motif
```

```
FT                   /note= "Claim 1"
FT   Peptide         329. .334
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         335. .340
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         354. .359
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         360. .365
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         366. .372
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         523. .528
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         542. .547
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         590. .595
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
XX
PN   US5643783-A.
XX
PD   01-JUL-1997.
XX
PF   01-DEC-1993;   93US-00159784.
XX
PR   01-DEC-1993;   93US-00159784.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Olsen BR,  Oh SP;
XX
DR   WPI; 1997-350247/32.
DR   N-PSDB; AAT84484.
XX
PT   Nucleic acid encoding human alpha-1 collagen - for production of
PT   recombinant alpha-1 collagen, for use in the treatment of cartilage
PT   degeneration.
XX
PS   Claim 1; Col 23-30; 35pp; English.
XX
CC   Novel human type alpha-1 (XVIII) collagen is characterised by 10 triple
CC   helical domains containing the GXYGX'Y' motif (where X, Y, X' and Y'
CC   represent any amino acid), the helical domains being separated and
CC   flanked by non-triple helical regions which may provide flexibility.
CC   Alpha-1 collagen is expressed in multiple tissues, especially liver, lung
CC   and kidney. A claimed plasmid comprising alpha-1 collagen nucleic acid
CC   (see AAT84484) and an expression control sequence can be used to express
CC   recombinant collagen in prokaryotic or eukaryotic (especially mammalian)
CC   host cells. The alpha-1 collagen may be used to treat a patient suffering
CC   from a disease associated with cartilage degradation, and for
CC   supplementing collagen. It can also be used as a connective tissue filler
CC   (e.g. for plastic surgery), can be interposed between a dermal equivalent

  Query Match             100.0%;  Score 893;  DB 2;  Length 684;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-100;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 573

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        574 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 633

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 683
```

```
RESULT 30
AAY25113
ID   AAY25113 standard; protein; 684 AA.
XX
AC   AAY25113;
XX
DT   25-AUG-1999  (first entry)
XX
DE   Human alpha1 (XVIII) collagen protein.
XX
KW   Alpha1(XVIII) collagen; mimetic; endostatin; atomic coordinate; library;
KW   anti-angiogenic; heparin binding domain; receptor binding domain; mimic;
KW   alpha-helix A domain; carbohydrate recognition domain; CRD domain;
KW   treatment; angiogenesis; tumour; human.
XX
OS   Homo sapiens.
XX
PN   WO9931616-A1.
XX
PD   24-JUN-1999.
XX
PF   16-DEC-1998;   98WO-US026783.
XX
PR   16-DEC-1997;   97US-0069727P.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Olsen BR,  Hohenester E,  Timpl R,  Sasaki T;
XX
DR   WPI; 1999-395243/33.
DR   N-PSDB; AAX78379.
XX
PT   Identifying mimetics of mammalian endostatin.
XX
PS   Disclosure; Fig 5A-C; 75pp; English.
XX
CC   This invention describes a novel method for identifying mimetics of
CC   mammalian endostatin. The method comprises identifying a compound having
CC   atomic coordinates with non-trivial similarity to selected coordinates of
CC   atoms of a mammalian endostatin involves (a) providing a library of
CC   atomic coordinates of compounds in a library of candidate compounds, (b)
CC   comparing the library of atomic coordinates to the selected coordinates
CC   of a mammalian endostatin and (c) selecting from the library at least one
CC   candidate compound on the basis of selection criteria which include
CC   similarities between the atomic coordinates of the selected candidate
CC   compound and the atomic coordinates of the mammalian endostatin. The
CC   invention also describes the use of an anti-angiogenic fragment of
CC   endostatin comprising a domain selected from a heparin binding domain, a
CC   receptor binding domain, and exposed on alpha-helix A domain, and a
CC   carbohydrate recognition domain (CRD) domain. The methods can be used for
CC   designing and selecting endostatin mimics. The compounds identified can
CC   be used for treating undesired angiogenesis, e.g. tumours. This sequence
CC   represents human alpha1(XVIII) collagen which is used in the description
CC   of the method
XX
SQ   Sequence 684 AA;

  Query Match             100.0%;  Score 893;  DB 2;  Length 684;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-100;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 573

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        574 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 633
```

```
Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 683
```

RESULT 31
AAO17357
ID AAO17357 standard; protein; 684 AA.
XX
AC AAO17357;
XX
DT 19-JUL-2002 (first entry)
XX
DE Human collagen type XVIII alpha 1.
XX
KW Human; endometriosis; DNA chip; fibronectin; p27; reticulocalbin;
KW aldehyde dehydrogenase 6; gravin; phospholipase C epsilon; elastin;
KW insulin-like growth factor binding protein-2; alpha-2 type IV collagen;
KW transmembrane receptor PTK7; collagen type XVIII alpha 1;
KW platelet derived growth factor receptor alpha; laminin M chain;
KW subtilisin like protein PACE4; nidogen.
XX
OS Homo sapiens.
XX
PN EP1191107-A2.
XX
PD 27-MAR-2002.
XX
PF 21-AUG-2001; 2001EP-00250300.
XX
PR 25-SEP-2000; 2000DE-01048633.
XX
PA (SCHD) SCHERING AG.
XX
PI Hess-Stumpp H, Haendler B, Kraetzschmar J, Kreft B, Winterhager E;
PI Regidor P, Scotti S;
XX
DR WPI; 2002-317413/36.
XX
PT In vitro diagnosis and monitoring of endometriosis, comprises detecting
PT reduced expression of specific gene products, e.g. from the fibronectin
PT gene.
XX
PS Claim 1; Page 12-13; 21pp; German.
XX
CC The present invention relates to a method for the in vitro diagnosis of
CC endometriosis by determining the amount of gene product from at least one
CC specific gene in a patient sample and comparing this with the amount of
CC gene product in a control sample. A reduced level is indicative of
CC endometriosis. The gene products may be fibronectin, p27, reticulocalbin,
CC aldehyde dehydrogenase 6, gravin, phospholipase C epsilon, elastin,
CC insulin-like growth factor binding protein-2, alpha-2 type IV collagen,
CC transmembrane receptor PTK7, collagen type XVIII alpha 1, platelet
CC derived growth factor receptor alpha, laminin M chain, subtilisin like
CC protein PACE4 or nidogen. The method is useful for initial diagnosis of
CC endometriosis, and also for monitoring progress and treatment of the
CC disease. The present sequence is human collagen type XVIII alpha 1
XX
SQ Sequence 684 AA;

```
  Query Match             100.0%;  Score 893;  DB 5;  Length 684;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-100;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 573

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        574 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 633

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 683
```

RESULT 32
AAW92296
ID AAW92296 standard; peptide; 1301 AA.
XX
AC AAW92296;
XX
DT 28-APR-1999 (first entry)
XX
DE Human alpha-1 (XVIII) collagen chain common sequence HU18(common)36.
XX
KW Human; type XVIII collagen; liver disease; cirrhosis; detection;
KW hepatocellular carcinoma; diagnosis.
XX
OS Homo sapiens.
XX
PN WO9856399-A1.
XX
PD 17-DEC-1998.
XX
PF 12-JUN-1998; 98WO-US012327.
XX
PR 12-JUN-1997; 97US-0049369P.
XX
PA (FIBR-) FIBROGEN INC.
PA (FIFI-) ACAD FINLAND.
PA (INRM) INST NAT SANTE & RECH MEDICALE.
XX
PI Pihlajaniemi T, Rehn M, Clement B;
XX
DR WPI; 1999-070292/06.
XX
PT Diagnosis and monitoring of liver disease by measuring collagen type
PT XVIII levels - with elevated levels indicative of disease, especially
PT cirrhosis or hepatocellular carcinoma.
XX
PS Example 6; Fig 8; 56pp; English.
XX
CC A method has been developed for the detecting liver disease. The method
CC comprises: (a) reacting a patient sample with antibodies (Ab) specific
CC for collagen type XVIII (Col18); (b) measuring the amount of Ab-antigen
CC complex (C) formed as indicator of the amount of Col18 present; (c)
CC similar analysis of a non-diseased control; and (d) comparing the amounts
CC of Col18 in the two samples to detect presence or progression of disease.
CC Elevated levels of Col18 are: (i) indicative of disease, specifically
CC cirrhosis; and (ii) predictive of the prognosis of disease, specifically
CC hepatocellular carcinoma (there is a relationship between Col18 mRNA
CC levels and tumour size and necrosis, and survival times are significantly
CC higher in patients with higher Col18 levels). The method provides non-
CC invasive, early and accurate diagnosis of liver disease. The present
CC sequence represents the sequence common to human alpha-1 (XVIII) collagen
CC chain from the present invention
XX
SQ Sequence 1301 AA;

```
  Query Match             100.0%;  Score 893;  DB 2;  Length 1301;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-100;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1131 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 1190

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1191 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 1250

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1251 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 1300
```

```
RESULT 33
AAY08694
ID   AAY08694 standard; protein; 1336 AA.
XX
AC   AAY08694;
XX
DT   10-AUG-1999  (first entry)
XX
DE   Human collagen 18 protein.
XX
KW   Plasminogen; human; angiostatin; endostatin; gene therapy; vector;
KW   anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
KW   tumour growth; solid tumour; diabetic retinopathy; retina; collagen 18.
XX
OS   Homo sapiens.
XX
PN   WO9926480-A1.
XX
PD   03-JUN-1999.
XX
PF   20-NOV-1998;   98WO-US024950.
XX
PR   20-NOV-1997;   97US-00975424.
XX
PA   (GENE-) GENETIX PHARM INC.
PA   (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX
PI   Leboulch P,  Pawliuk RJ,  Bachelot T;
XX
DR   WPI; 1999-357696/30.
DR   N-PSDB; AAX77720.
XX
PT   Anti-angiogenic gene therapy vectors.
XX
PS   Disclosure; Page 77-80; 83pp; English.
XX
CC   This invention describes a novel viral gene therapy vector comprising a
CC   nucleic acid molecule encoding an anti-angiogenic polypeptide chosen from
CC   human or murine angiostatin, human or murine endostatin and angiogenesis-
CC   inhibiting fusions and fragments, where the viral vector is sufficiently
CC   attenuated for use in human gene therapy. The products of the invention
CC   have anti-angiogenic, cytostatic, anti-diabetic and ophthalmological
CC   activity. The vector is used in gene therapy for inhibiting tumour growth
CC   in humans harbouring a solid tumour. The vector expresses an anti-
CC   angiogenic polypeptide. An additional use comprises treatment of diabetic
CC   retinopathy, where the anti-angiogenic polypeptide inhibits angiogenesis
CC   in the vicinity of the retina. The vector is administered to cells ex
CC   vivo and then administered to the patient
XX
SQ   Sequence 1336 AA;

  Query Match             100.0%;  Score 893;  DB 2;  Length 1336;
  Best Local Similarity   100.0%;  Pred. No. 6e-100;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1166 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 1225

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1226 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 1285

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1286 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 1335


RESULT 34
ABP96308
ID   ABP96308 standard; protein; 1336 AA.
XX
AC   ABP96308;
XX
DT   20-MAY-2003  (first entry)
XX
DE   Human endostatin protein.
XX
KW   Humanised baculovirus; cytostatic; gene therapy; baculovirus; cancer;
KW   prostate cancer; endostatin; chromosome 21.
XX
OS   Homo sapiens.
XX
PN   WO2003016540-A2.
XX
PD   27-FEB-2003.
XX
PF   15-AUG-2002; 2002WO-GB003791.
XX
PR   15-AUG-2001; 2001GB-00019852.
XX
PA   (UYYO-) UNIV YORK.
XX
PI   Maitland N;
XX
DR   WPI; 2003-268336/26.
XX
PT   New baculovirus having a modified genome encoding a therapeutic agent,
PT   useful in the manufacture of a medicament for the treatment of cancer,
PT   particularly prostate cancer.
XX
PS   Claim 24; Page; 34pp; English.
XX
CC   The present invention describes a humanised baculovirus (I) which
CC   comprises a modified baculovirus genome having a nucleic acid molecule
CC   encoding a therapeutic agent and a polypeptide which functions to target
CC   the baculovirus to at least one cell type. Also described is a
CC   pharmaceutical composition comprising (I). (I) has cytostatic activity,
CC   and can be used in gene therapy. The baculovirus is useful in the
CC   manufacture of a medicament for the treatment of cancer, particularly
CC   prostate cancer. The present sequence represents the human endostatin
CC   protein, which is specified in the exemplification of the present
CC   invention. N.B. The present sequence is not given in the specification
CC   but is referred to in Claim 24 as Genbank accession number NM_130445
XX
SQ   Sequence 1336 AA;

  Query Match             100.0%;  Score 893;  DB 6;  Length 1336;
  Best Local Similarity   100.0%;  Pred. No. 6e-100;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1166 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 1225

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1226 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 1285

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1286 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 1335


RESULT 35
ABB83471
ID   ABB83471 standard; protein; 1516 AA.
XX
AC   ABB83471;
XX
DT   30-SEP-2002  (first entry)
XX
DE   Human collagen XVIII.
XX
```

```
KW   Human; antirheumatic; antiarthritic; gene therapy; anti-angiogenic;
KW   rheumatoid arthritis; collagen; endostatin.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Protein         1334. .1516
FT                   /note= "This region is specifically claimed in Claim 4"
XX
PN   WO200253191-A1.
XX
PD   11-JUL-2002.
XX
PF   03-JAN-2002; 2002WO-KR000001.
XX
PR   05-JAN-2001; 2001KR-00000691.
XX
PA   (VIRO-) VIROMED LTD.
XX
PI   Kim J,  Ho S,  Park E,  Kim S;
XX
DR   WPI; 2002-583596/62.
DR   N-PSDB; ABN85301.
XX
PT   Novel composition for gene therapy against rheumatoid arthritis,
PT   comprising a DNA encoding anti-angiogenic protein or its parts.
XX
PS   Claim 4; Page 70-78; 84pp; English.
XX
CC   The present invention relates to a composition for gene therapy,
CC   comprising a DNA encoding an anti-angiogenic protein, which shows
CC   therapeutic effects on rheumatoid arthritis. The composition is useful
CC   for treating rheumatoid arthritis and the gene therapy is effective,
CC   lasting for 14 days. The present sequence is the protein sequence for
CC   human collagen XVIII. Endostatin, which consists of the C-terminal 183
CC   residues of collagen XVIII, was used as an anti-angiogenic protein
XX
SQ   Sequence 1516 AA;

  Query Match             100.0%;  Score 893;  DB 5;  Length 1516;
  Best Local Similarity   100.0%;  Pred. No. 7.2e-100;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1346 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 1405

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1406 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 1465

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1466 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 1515


RESULT 36
ABP68617
ID   ABP68617 standard; protein; 1516 AA.
XX
AC   ABP68617;
XX
DT   14-JAN-2003  (first entry)
XX
DE   Human pancreatic cancer expressed protein SEQ ID NO 166.
XX
KW   Human; pancreas; cancer; gene therapy; vaccine; immunostimulant;
KW   cytostatic; tumour.
XX
OS   Homo sapiens.
XX
PN   WO200260317-A2.
XX
PD   08-AUG-2002.
XX
PF   30-JAN-2002; 2002WO-US002781.
XX
PR   30-JAN-2001; 2001US-0265305P.
PR   31-JAN-2001; 2001US-0265682P.
PR   09-FEB-2001; 2001US-0267568P.
PR   21-MAR-2001; 2001US-0278651P.
PR   28-APR-2001; 2001US-0287112P.
PR   16-MAY-2001; 2001US-0291631P.
PR   12-JUL-2001; 2001US-0305484P.
PR   20-AUG-2001; 2001US-0313999P.
PR   27-NOV-2001; 2001US-0333626P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Benson DR,  Kalos MD,  Lodes MJ,  Persing DH,  Hepler WT,  Jiang Y;
XX
DR   WPI; 2002-627435/67.
DR   N-PSDB; ABV94763.
XX
PT   New isolated polynucleotide and pancreatic tumor polypeptides, useful for
PT   diagnosing, preventing and/or treating cancer, particularly pancreatic
PT   cancer.
XX
PS   Claim 2; SEQ ID NO 166; 300pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated polynucleotide (I) comprising: (a)
CC   any of a group of over 4000 nucleotide sequences (ABV94628-ABV99145); (b)
CC   complements of (a); (c) sequences consisting of at least 20 contiguous
CC   residues of (a); (d) sequences that hybridize to (a), under moderately
CC   stringent conditions; (e) sequences having at least 75% or 90% identity
CC   to (a); or (f) degenerate variants of (a). Polypeptides (ABP68596-
CC   ABP68637) encoded by (I) and oligonucleotide can be used to detect cancer
CC   in a patient and compositions comprising polypeptides, polynucleotides,
CC   antibodies, fusion proteins, T cell populations and antigen presenting
CC   cells expressing the polypeptide are useful in treating pancreatic cancer
CC   and stimulating an immune response. The polynucleotides can be used as
CC   probes or primers for nucleic acid hybridisation, in the design and
CC   preparation of ribozyme molecules for inhibiting expression of the tumour
CC   polypeptides and proteins in the tumour cells, in vaccines and for gene
CC   therapy. Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1516 AA;

  Query Match             100.0%;  Score 893;  DB 5;  Length 1516;
  Best Local Similarity   100.0%;  Pred. No. 7.2e-100;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1346 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 1405

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1406 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 1465

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1466 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 1515


Search completed: March 13, 2004, 08:29:11
Job time : 59 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 13, 2004, 08:17:57 ; Search time 34 Seconds
(without alignments)
1055.766 Million cell updates/sec

Title: US-09-171-607A-1
Perfect score: 893
Sequence: 1 VALNSPLSGGMRGIRGADFQ.....ASCHHAYIVLCIENSFMTAS 170

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	893	100.0	178	14	US-10-131-241-60	Sequence 60, Appl
2	893	100.0	178	14	US-10-042-347-5	Sequence 5, Appli
3	893	100.0	179	14	US-10-131-241-57	Sequence 57, Appl
4	893	100.0	182	9	US-09-998-831-14	Sequence 14, Appl
5	893	100.0	182	14	US-10-131-241-54	Sequence 54, Appl
6	893	100.0	182	14	US-10-042-347-3	Sequence 3, Appli
7	893	100.0	182	14	US-10-373-561-14	Sequence 14, Appl
8	893	100.0	183	9	US-09-873-676-2	Sequence 2, Appli
9	893	100.0	183	13	US-10-080-797-1	Sequence 1, Appli
10	893	100.0	183	14	US-10-131-241-52	Sequence 52, Appl
11	893	100.0	183	14	US-10-292-418-4	Sequence 4, Appli
12	893	100.0	682	15	US-10-264-049-3010	Sequence 3010, Ap
13	893	100.0	684	10	US-09-961-403-5	Sequence 5, Appli
14	893	100.0	1516	14	US-10-060-036-166	Sequence 166, App
15	893	100.0	1516	15	US-10-431-642-3	Sequence 3, Appli
16	889	99.6	181	14	US-10-131-241-55	Sequence 55, Appl
17	885	99.1	180	14	US-10-131-241-56	Sequence 56, Appl
18	858	96.1	180	14	US-10-131-241-47	Sequence 47, Appl
19	783	87.7	184	10	US-09-938-391-4	Sequence 4, Appli
20	783	87.7	184	14	US-10-131-241-49	Sequence 49, Appl
21	783	87.7	184	14	US-10-292-418-35	Sequence 35, Appl
22	783	87.7	230	10	US-09-938-391-2	Sequence 2, Appli
23	778	87.1	191	9	US-09-998-831-13	Sequence 13, Appl
24	778	87.1	191	14	US-10-373-561-13	Sequence 13, Appl
25	778	87.1	207	13	US-10-080-797-3	Sequence 3, Appli
26	775	86.8	184	14	US-10-131-241-46	Sequence 46, Appl
27	775	86.8	184	14	US-10-292-418-18	Sequence 18, Appl
28	775	86.8	207	14	US-10-422-934-71	Sequence 71, Appl
29	734	82.2	185	13	US-10-036-869-36	Sequence 36, Appl
30	346	38.7	63	9	US-09-822-540A-1	Sequence 1, Appli
31	182	20.4	31	9	US-09-822-540A-2	Sequence 2, Appli
32	109	12.2	24	15	US-10-016-569A-29	Sequence 29, Appl
33	109	12.2	24	15	US-10-308-644-29	Sequence 29, Appl
34	94	10.5	16	9	US-09-766-412-32	Sequence 32, Appl
35	85	9.5	1244	9	US-09-815-915-8	Sequence 8, Appli
36	85	9.5	1244	14	US-10-393-316-8	Sequence 8, Appli
37	80	9.0	409	14	US-10-156-761-13624	Sequence 13624, A
38	78.5	8.8	285	15	US-10-369-493-12038	Sequence 12038, A
39	77.5	8.7	418	9	US-09-927-602-5	Sequence 5, Appli
40	77	8.6	451	15	US-10-369-493-9525	Sequence 9525, Ap
41	77	8.6	1267	10	US-09-769-734-38	Sequence 38, Appl
42	76.5	8.6	421	14	US-10-156-761-13612	Sequence 13612, A
43	76	8.5	239	15	US-10-369-493-8031	Sequence 8031, Ap
44	76	8.5	636	15	US-10-351-157-9	Sequence 9, Appli
45	75.5	8.5	3192	14	US-10-132-134-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-10-131-241-60
; Sequence 60, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
; APPLICANT: Holaday, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolife
; TITLE OF INVENTION: and Regulating Angiogenesis Using Cancer Markers
; FILE REFERENCE: 05213-0344 43170-271565
; CURRENT APPLICATION NUMBER: US/10/131,241
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/413,049
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: US 09/316,802
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US 60/086,586
; PRIOR FILING DATE: 1998-05-22
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-241-60

Query Match 100.0%; Score 893; DB 14; Length 178;
Best Local Similarity 100.0%; Pred. No. 3.4e-92;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          9 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 68

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 128
```

```
Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 178
```

RESULT 2
US-10-042-347-5
; Sequence 5, Application US/10042347
; Publication No. US20030114370A1
; GENERAL INFORMATION:
; APPLICANT: O'Reilly, Michael S.
; APPLICANT: Folkman, M. Judah
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Endostatin Protein and Peptide Fr
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 05213-0880 (43170-249874)
; CURRENT APPLICATION NUMBER: US/10/042,347
; CURRENT FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: US 09/315,689
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: US 60/106,343
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: US 09/154,302
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: US 08/740,168
; PRIOR FILING DATE: 1996-10-22
; PRIOR APPLICATION NUMBER: US 60/005,835
; PRIOR FILING DATE: 1995-10-23
; PRIOR APPLICATION NUMBER: US 60/023,070
; PRIOR FILING DATE: 1996-08-02
; PRIOR APPLICATION NUMBER: US 60/026,263
; PRIOR FILING DATE: 1996-09-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-042-347-5

Query Match 100.0%; Score 893; DB 14; Length 178;
Best Local Similarity 100.0%; Pred. No. 3.4e-92;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          9 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 68

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 128

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 178
```

RESULT 3
US-10-131-241-57
; Sequence 57, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
; APPLICANT: Holaday, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolifer
; TITLE OF INVENTION: and Regulating Angiogenesis Using Cancer Markers
; FILE REFERENCE: 05213-0344 43170-271565
; CURRENT APPLICATION NUMBER: US/10/131,241
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/413,049
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: US 09/316,802
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US 60/086,586
; PRIOR FILING DATE: 1998-05-22
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-241-57

Query Match 100.0%; Score 893; DB 14; Length 179;
Best Local Similarity 100.0%; Pred. No. 3.4e-92;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          9 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 68

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 128

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 178
```

RESULT 4
US-09-998-831-14
; Sequence 14, Application US/09998831
; Patent No. US20020119153A1
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY
; TITLE OF INVENTION: INHIBITING VEGF
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/998,831
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 09/561,108
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: PEPTIDE
US-09-998-831-14

Query Match 100.0%; Score 893; DB 9; Length 182;
Best Local Similarity 100.0%; Pred. No. 3.5e-92;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182
```

RESULT 5
US-10-131-241-54
; Sequence 54, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:

; APPLICANT: Holaday, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolifer
; TITLE OF INVENTION: and Regulating Angiogenesis Using Cancer Markers
; FILE REFERENCE: 05213-0344 43170-271565
; CURRENT APPLICATION NUMBER: US/10/131,241
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/413,049
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: US 09/316,802
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US 60/086,586
; PRIOR FILING DATE: 1998-05-22
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-241-54

Query Match 100.0%; Score 893; DB 14; Length 182;
Best Local Similarity 100.0%; Pred. No. 3.5e-92;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182
```

RESULT 6
US-10-042-347-3
; Sequence 3, Application US/10042347
; Publication No. US20030114370A1
; GENERAL INFORMATION:
; APPLICANT: O'Reilly, Michael S.
; APPLICANT: Folkman, M. Judah
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Endostatin Protein and Peptide Fr
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 05213-0880 (43170-249874)
; CURRENT APPLICATION NUMBER: US/10/042,347
; CURRENT FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: US 09/315,689
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: US 60/106,343
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: US 09/154,302
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: US 08/740,168
; PRIOR FILING DATE: 1996-10-22
; PRIOR APPLICATION NUMBER: US 60/005,835
; PRIOR FILING DATE: 1995-10-23
; PRIOR APPLICATION NUMBER: US 60/023,070
; PRIOR FILING DATE: 1996-08-02
; PRIOR APPLICATION NUMBER: US 60/026,263
; PRIOR FILING DATE: 1996-09-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-042-347-3

Query Match 100.0%; Score 893; DB 14; Length 182;
Best Local Similarity 100.0%; Pred. No. 3.5e-92;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182
```

RESULT 7
US-10-373-561-14
; Sequence 14, Application US/10373561
; Publication No. US20030175276A1
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002582
; CURRENT APPLICATION NUMBER: US/10/373,561
; CURRENT FILING DATE: 2003-02-24
; PRIOR APPLICATION NUMBER: US/09/561,499
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: PEPTIDE
US-10-373-561-14

Query Match 100.0%; Score 893; DB 14; Length 182;
Best Local Similarity 100.0%; Pred. No. 3.5e-92;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182
```

RESULT 8
US-09-873-676-2
; Sequence 2, Application US/09873676
; Patent No. US20020077289A1
; GENERAL INFORMATION:
; APPLICANT: MacDonald, Nicholas J.
; APPLICANT: Sim, Kim L.
; TITLE OF INVENTION: Angiostatin and Endostatin Binding Proteins and Methods of Use
; FILE REFERENCE: 05213-0378 (43170-259333)
; CURRENT APPLICATION NUMBER: US/09/873,676
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: US 60/209,065
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: US 60/289,387

```
;  PRIOR FILING DATE: 2001-05-08
;  NUMBER OF SEQ ID NOS: 123
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;    LENGTH: 183
;    TYPE: PRT
;    ORGANISM: Homo sapiens
US-09-873-676-2

  Query Match             100.0%;  Score 893;  DB 9;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-92;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 9
US-10-080-797-1
; Sequence 1, Application US/10080797
; Publication No. US20020183253A1
; GENERAL INFORMATION:
;  APPLICANT: Campochiaro, Peter A.
;  APPLICANT:  Dixon, Katharine H.
;  APPLICANT:  Brazzell, Romulus K.
;  TITLE OF INVENTION: METHOD FOR TREATING OCULAR
;  TITLE OF INVENTION:  NEOVASCULARIZATION
;  FILE REFERENCE: 4-31881A
;  CURRENT APPLICATION NUMBER: US/10/080,797
;  CURRENT FILING DATE:  2002-02-21
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;    LENGTH: 183
;    TYPE: PRT
;    ORGANISM: Human
US-10-080-797-1

  Query Match             100.0%;  Score 893;  DB 13;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-92;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 10
US-10-131-241-52
; Sequence 52, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
;  APPLICANT: Holaday, John W.
;  APPLICANT:  Fortier, Anne H.
;  TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolifer
;  TITLE OF INVENTION:  and Regulating Angiogenesis Using Cancer Markers
;  FILE REFERENCE: 05213-0344  43170-271565
;  CURRENT APPLICATION NUMBER: US/10/131,241
;  CURRENT FILING DATE:  2002-07-22
;  PRIOR APPLICATION NUMBER: US 09/413,049
;  PRIOR FILING DATE: 1999-10-06
;  PRIOR APPLICATION NUMBER: US 09/316,802
;  PRIOR FILING DATE: 1999-05-21
;  PRIOR APPLICATION NUMBER: US 60/086,586
;  PRIOR FILING DATE: 1998-05-22
;  NUMBER OF SEQ ID NOS: 65
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52
;    LENGTH: 183
;    TYPE: PRT
;    ORGANISM: Homo sapiens
US-10-131-241-52

  Query Match             100.0%;  Score 893;  DB 14;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-92;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 11
US-10-292-418-4
; Sequence 4, Application US/10292418
; Publication No. US20030139365A1
; GENERAL INFORMATION:
;  APPLICANT: Lo, Kin-Ming
;  APPLICANT:  Li, Yue
;  APPLICANT:  Gillies, Stephen D
;  TITLE OF INVENTION: Expression and Export of Angiogenesis Inhibitors as
;  TITLE OF INVENTION:  Immunofusins
;  FILE REFERENCE: LEX-006C1
;  CURRENT APPLICATION NUMBER: US/10/292,418
;  CURRENT FILING DATE:  2002-11-12
;  PRIOR APPLICATION NUMBER: 09/383,315
;  PRIOR FILING DATE: 1999-08-25
;  PRIOR APPLICATION NUMBER: US 60/097,883
;  PRIOR FILING DATE: 1998-08-25
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;    LENGTH: 183
;    TYPE: PRT
;    ORGANISM: Homo sapiens
US-10-292-418-4

  Query Match             100.0%;  Score 893;  DB 14;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-92;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 12
US-10-264-049-3010
; Sequence 3010, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
;  APPLICANT: Birse et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA133P1
;  CURRENT APPLICATION NUMBER: US/10/264,049
;  CURRENT FILING DATE:  2002-10-04
;  PRIOR APPLICATION NUMBER: PCT/US01/18569
;  PRIOR FILING DATE: 2001-06-07
;  PRIOR APPLICATION NUMBER: US 60/209,467
;  PRIOR FILING DATE: 2000-06-07
;  NUMBER OF SEQ ID NOS: 4360
;  SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 3010
;   LENGTH: 682
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (20)
;   OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (39)
;   OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-3010

  Query Match             100.0%;  Score 893;  DB 15;  Length 682;
  Best Local Similarity   100.0%;  Pred. No. 2e-91;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        512 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 571

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        572 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 631

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        632 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 681


RESULT 13
US-09-961-403-5
; Sequence 5, Application US/09961403
; Publication No. US20030077589A1
; GENERAL INFORMATION:
;  APPLICANT: HE-STUMPP, HOLGER
;  APPLICANT:  HAENDLER, BERNARD
;  APPLICANT:  KRAETZSCHMAR, JOERN
;  APPLICANT:  KREFT, BERTHOLT
;  APPLICANT:  WINTERHAGER, ELKE
;  APPLICANT:  REGIDOR, PEDRO
;  APPLICANT:  SCOTTI, SIMONE
;  TITLE OF INVENTION: METHOD FOR IN VITRO DIAGNOSIS OF ENDOMETRIOSIS
;  FILE REFERENCE: SCH-1789
;  CURRENT APPLICATION NUMBER: US/09/961,403
;  CURRENT FILING DATE:  2001-09-25
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 684
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-961-403-5

  Query Match             100.0%;  Score 893;  DB 10;  Length 684;
  Best Local Similarity   100.0%;  Pred. No. 2e-91;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 573

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        574 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 633

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 683


RESULT 14
US-10-060-036-166
; Sequence 166, Application US/10060036
; Publication No. US20030073144A1
; GENERAL INFORMATION:
;  APPLICANT: Benson, Darin R.
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Persing, David H.
;  APPLICANT:  Hepler, William T.
;  APPLICANT:  Jiang, Yuqiu
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF PANCREATIC CANCER
;  FILE REFERENCE: 210121.566
;  CURRENT APPLICATION NUMBER: US/10/060,036
;  CURRENT FILING DATE:  2002-01-30
;  NUMBER OF SEQ ID NOS: 4560
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 166
;   LENGTH: 1516
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-060-036-166

  Query Match             100.0%;  Score 893;  DB 14;  Length 1516;
  Best Local Similarity   100.0%;  Pred. No. 5.6e-91;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1346 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 1405

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1406 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 1465

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1466 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 1515


RESULT 15
US-10-431-642-3
; Sequence 3, Application US/10431642
; Publication No. US20040009920A1
; GENERAL INFORMATION:
;  APPLICANT: Ruoslahti, Erkki
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING
;  TITLE OF INVENTION:  TUMOR GROWTH AND ANGIOGENESIS
;  FILE REFERENCE: BURNHAM.008CIP
;  CURRENT APPLICATION NUMBER: US/10/431,642
;  CURRENT FILING DATE:  2003-05-05
;  PRIOR APPLICATION NUMBER: 10/005,171
;  PRIOR FILING DATE: 2001-12-03
;  PRIOR APPLICATION NUMBER: 60/331,357
```

; PRIOR FILING DATE: 2000-12-04
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1516
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-431-642-3

Query Match 100.0%; Score 893; DB 15; Length 1516;
Best Local Similarity 100.0%; Pred. No. 5.6e-91;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1346 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 1405

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1406 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 1465

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1466 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 1515
```

Search completed: March 13, 2004, 08:23:27
Job time : 35 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 13, 2004, 08:15:22 ; Search time 22 Seconds
(without alignments)
398.928 Million cell updates/sec

Title: US-09-171-607A-1
Perfect score: 893
Sequence: 1 VALNSPLSGGMRGIRGADFQ.....ASCHHAYIVLCIENSFMTAS 170

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	893	100.0	178	4	US-09-315-689-5	Sequence 5, Appli
2	893	100.0	182	4	US-09-561-500-14	Sequence 14, Appl
3	893	100.0	182	4	US-09-561-108-14	Sequence 14, Appl
4	893	100.0	182	4	US-09-315-689-3	Sequence 3, Appli
5	893	100.0	182	4	US-09-561-526-14	Sequence 14, Appl
6	893	100.0	182	4	US-09-561-499-14	Sequence 14, Appl
7	893	100.0	182	4	US-09-998-831-14	Sequence 14, Appl
8	893	100.0	183	3	US-09-206-059-2	Sequence 2, Appli
9	778	87.1	191	4	US-09-561-500-13	Sequence 13, Appl
10	778	87.1	191	4	US-09-561-108-13	Sequence 13, Appl
11	778	87.1	191	4	US-09-561-526-13	Sequence 13, Appl
12	778	87.1	191	4	US-09-561-499-13	Sequence 13, Appl
13	778	87.1	191	4	US-09-998-831-13	Sequence 13, Appl
14	775	86.8	195	1	US-08-159-784-2	Sequence 2, Appli
15	734	82.2	185	3	US-08-985-526-36	Sequence 36, Appl
16	491	55.0	191	1	US-08-159-784-3	Sequence 3, Appli
17	242.5	27.2	123	4	US-09-231-077D-11	Sequence 11, Appl
18	214.5	24.0	124	4	US-09-231-077D-10	Sequence 10, Appl
19	160	17.9	35	3	US-09-046-985-2	Sequence 2, Appli
20	160	17.9	35	3	US-09-474-743-2	Sequence 2, Appli
21	101	11.3	22	3	US-09-046-985-7	Sequence 7, Appli
22	101	11.3	22	3	US-09-474-743-7	Sequence 7, Appli
23	94	10.5	16	3	US-09-385-442-32	Sequence 32, Appl
24	76	8.5	403	4	US-09-252-991A-22238	Sequence 22238, A
25	76	8.5	578	1	US-08-653-740-3	Sequence 3, Appli
26	76	8.5	578	2	US-09-073-594-3	Sequence 3, Appli
27	76	8.5	578	3	US-09-275-925-3	Sequence 3, Appli
28	76	8.5	636	1	US-08-653-740-5	Sequence 5, Appli
29	76	8.5	636	2	US-09-073-594-5	Sequence 5, Appli
30	76	8.5	636	3	US-09-275-925-5	Sequence 5, Appli
31	74	8.3	256	1	US-07-906-349A-8	Sequence 8, Appli
32	74	8.3	256	1	US-08-167-035-4	Sequence 4, Appli
33	74	8.3	256	1	US-08-208-887A-4	Sequence 4, Appli
34	74	8.3	256	2	US-08-539-005-4	Sequence 4, Appli
35	74	8.3	256	4	US-09-280-598-8	Sequence 8, Appli
36	73	8.2	311	3	US-08-987-743-2	Sequence 2, Appli
37	73	8.2	435	3	US-08-733-360A-1	Sequence 1, Appli
38	73	8.2	435	3	US-08-733-360A-3	Sequence 3, Appli
39	73	8.2	435	3	US-08-987-743-6	Sequence 6, Appli
40	73	8.2	435	3	US-08-987-743-15	Sequence 15, Appl
41	73	8.2	435	3	US-08-916-935-1	Sequence 1, Appli
42	73	8.2	435	3	US-08-916-935-3	Sequence 3, Appli
43	72.5	8.1	184	4	US-09-252-991A-19779	Sequence 19779, A
44	72	8.1	304	4	US-08-630-915A-28	Sequence 28, Appl
45	72	8.1	1006	4	US-09-023-905A-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-09-315-689-5
; Sequence 5, Application US/09315689
; Patent No. 6346510
; GENERAL INFORMATION:
; APPLICANT: Folkman, Judah
; APPLICANT: O'Reilly, Michael
; TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions
; FILE REFERENCE: 05213-0229
; CURRENT APPLICATION NUMBER: US/09/315,689
; CURRENT FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-315-689-5

Query Match 100.0%; Score 893; DB 4; Length 178;
Best Local Similarity 100.0%; Pred. No. 4.5e-101;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          9 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 68

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 128

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 178
```

RESULT 2
US-09-561-500-14
; Sequence 14, Application US/09561500
; Patent No. 6342219
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002500
; CURRENT APPLICATION NUMBER: US/09/561,500
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44

```
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 14
;   LENGTH: 182
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  SYNTHETIC
;   OTHER INFORMATION: PEPTIDE
US-09-561-500-14

  Query Match             100.0%;  Score 893;  DB 4;  Length 182;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 3
US-09-561-108-14
; Sequence 14, Application US/09561108
; Patent No. 6342221
; GENERAL INFORMATION:
;  APPLICANT:  Philip E. Thorpe
;  APPLICANT:  Rolf A. Brekken
;  TITLE OF INVENTION:  ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
;  FILE REFERENCE:  4001.002584
;  CURRENT APPLICATION NUMBER: US/09/561,108
;  CURRENT FILING DATE:  2000-04-28
;  PRIOR APPLICATION NUMBER:  60/131,432
;  PRIOR FILING DATE:  1999-04-28
;  NUMBER OF SEQ ID NOS:  44
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 14
;   LENGTH: 182
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  SYNTHETIC
;   OTHER INFORMATION: PEPTIDE
US-09-561-108-14

  Query Match             100.0%;  Score 893;  DB 4;  Length 182;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 4
US-09-315-689-3
; Sequence 3, Application US/09315689
; Patent No. 6346510
; GENERAL INFORMATION:
;  APPLICANT: Folkman, Judah
;  APPLICANT: O'Reilly, Michael
;  TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions
;  FILE REFERENCE: 05213-0229
;  CURRENT APPLICATION NUMBER: US/09/315,689
;  CURRENT FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 182
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-315-689-3

  Query Match             100.0%;  Score 893;  DB 4;  Length 182;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 5
US-09-561-526-14
; Sequence 14, Application US/09561526
; Patent No. 6416758
; GENERAL INFORMATION:
;  APPLICANT:  Philip E. Thorpe
;  APPLICANT:  Rolf A. Brekken
;  TITLE OF INVENTION:  ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF
;  FILE REFERENCE:  4001.002586
;  CURRENT APPLICATION NUMBER: US/09/561,526
;  CURRENT FILING DATE:  2000-04-28
;  PRIOR APPLICATION NUMBER:  60/131,432
;  PRIOR FILING DATE:  1999-04-28
;  NUMBER OF SEQ ID NOS:  44
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 14
;   LENGTH: 182
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  SYNTHETIC
;   OTHER INFORMATION: PEPTIDE
US-09-561-526-14

  Query Match             100.0%;  Score 893;  DB 4;  Length 182;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 6
```

```
US-09-561-499-14
; Sequence 14, Application US/09561499
; Patent No. 6524583
; GENERAL INFORMATION:
;  APPLICANT:  Philip E. Thorpe
;  APPLICANT:  Rolf A. Brekken
;  TITLE OF INVENTION:  ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
;  FILE REFERENCE:  4001.002582
;  CURRENT APPLICATION NUMBER: US/09/561,499
;  CURRENT FILING DATE:  2000-04-28
;  PRIOR APPLICATION NUMBER:  60/131,432
;  PRIOR FILING DATE:  1999-04-28
;  NUMBER OF SEQ ID NOS:  44
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 14
;   LENGTH: 182
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  SYNTHETIC
;   OTHER INFORMATION: PEPTIDE
US-09-561-499-14

  Query Match            100.0%;  Score 893;  DB 4;  Length 182;
  Best Local Similarity  100.0%;  Pred. No. 4.6e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 7
US-09-998-831-14
; Sequence 14, Application US/09998831
; Patent No. 6676941
; GENERAL INFORMATION:
;  APPLICANT: Philip E. Thorpe
;  APPLICANT:  Rolf A. Brekken
;  TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY
;  TITLE OF INVENTION:  INHIBITING VEGF
;  FILE REFERENCE: 4001.002584
;  CURRENT APPLICATION NUMBER: US/09/998,831
;  CURRENT FILING DATE:  2001-11-30
;  PRIOR APPLICATION NUMBER: 09/561,108
;  PRIOR FILING DATE: 2000-04-28
;  NUMBER OF SEQ ID NOS: 44
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
;   LENGTH: 182
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  SYNTHETIC
;   OTHER INFORMATION: PEPTIDE
US-09-998-831-14

  Query Match            100.0%;  Score 893;  DB 4;  Length 182;
  Best Local Similarity  100.0%;  Pred. No. 4.6e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 8
US-09-206-059-2
; Sequence 2, Application US/09206059
; Patent No. 6201104
; GENERAL INFORMATION:
;  APPLICANT: MacDonald, Nicholas
;  APPLICANT: Sim, Kim Lee
;  TITLE OF INVENTION: Angiogenesis-Inhibiting Protein Binding Peptides and
;  TITLE OF INVENTION: Proteins and Methods of Use
;  FILE REFERENCE: 05213-0370
;  CURRENT APPLICATION NUMBER: US/09/206,059
;  CURRENT FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 183
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-206-059-2

  Query Match            100.0%;  Score 893;  DB 3;  Length 183;
  Best Local Similarity  100.0%;  Pred. No. 4.7e-101;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 72

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 132

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182


RESULT 9
US-09-561-500-13
; Sequence 13, Application US/09561500
; Patent No. 6342219
; GENERAL INFORMATION:
;  APPLICANT:  Philip E. Thorpe
;  APPLICANT:  Rolf A. Brekken
;  TITLE OF INVENTION:  ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
;  FILE REFERENCE:  4001.002500
;  CURRENT APPLICATION NUMBER: US/09/561,500
;  CURRENT FILING DATE:  2000-04-28
;  PRIOR APPLICATION NUMBER:  60/131,432
;  PRIOR FILING DATE:  1999-04-28
;  NUMBER OF SEQ ID NOS:  44
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 191
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  SYNTHETIC
US-09-561-500-13

  Query Match            87.1%;  Score 778;  DB 4;  Length 191;
  Best Local Similarity  85.8%;  Pred. No. 5.6e-87;
  Matches  145;  Conservative   13;  Mismatches   11;  Indels    0;  Gaps    0;
```

```
Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||:|||||||||||||||||||||||||||:||||||||||||||||||||| :||||
Db         20 VALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRGSVPIV 79

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              |||||:| |||::|||||:| |:||||||||||:|||||| ||||||||||||:|||| |
Db         80 NLKDEVLSPSWDSLFSGSQGQLQPGARIFSFDGRDVLRHPAWPQKSVWHGSDPSGRRLME 139

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 169
              |||||||||   |||||||| |||| | |||||::||||||||||||:
Db        140 SYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMTS 188


RESULT 10
US-09-561-108-13
; Sequence 13, Application US/09561108
; Patent No. 6342221
; GENERAL INFORMATION:
;  APPLICANT:  Philip E. Thorpe
;  APPLICANT:  Rolf A. Brekken
;  TITLE OF INVENTION:  ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
;  FILE REFERENCE:  4001.002584
;  CURRENT APPLICATION NUMBER: US/09/561,108
;  CURRENT FILING DATE:  2000-04-28
;  PRIOR APPLICATION NUMBER:  60/131,432
;  PRIOR FILING DATE:  1999-04-28
;  NUMBER OF SEQ ID NOS:  44
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 191
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  SYNTHETIC
US-09-561-108-13

  Query Match             87.1%;  Score 778;  DB 4;  Length 191;
  Best Local Similarity   85.8%;  Pred. No. 5.6e-87;
  Matches  145;  Conservative   13;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||:|||||||||||||||||||||||||||:||||||||||||||||||||| :||||
Db         20 VALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRGSVPIV 79

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              |||||:| |||::|||||:| |:||||||||||:|||||| ||||||||||||:|||| |
Db         80 NLKDEVLSPSWDSLFSGSQGQLQPGARIFSFDGRDVLRHPAWPQKSVWHGSDPSGRRLME 139

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 169
              |||||||||   |||||||| |||| | |||||::||||||||||||:
Db        140 SYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMTS 188


RESULT 11
US-09-561-526-13
; Sequence 13, Application US/09561526
; Patent No. 6416758
; GENERAL INFORMATION:
;  APPLICANT:  Philip E. Thorpe
;  APPLICANT:  Rolf A. Brekken
;  TITLE OF INVENTION:  ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF
;  FILE REFERENCE:  4001.002586
;  CURRENT APPLICATION NUMBER: US/09/561,526
;  CURRENT FILING DATE:  2000-04-28
;  PRIOR APPLICATION NUMBER:  60/131,432
;  PRIOR FILING DATE:  1999-04-28
;  NUMBER OF SEQ ID NOS:  44
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 191
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  SYNTHETIC
US-09-561-526-13

  Query Match             87.1%;  Score 778;  DB 4;  Length 191;
  Best Local Similarity   85.8%;  Pred. No. 5.6e-87;
  Matches  145;  Conservative   13;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||:|||||||||||||||||||||||||||:||||||||||||||||||||| :||||
Db         20 VALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRGSVPIV 79

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              |||||:| |||::|||||:| |:||||||||||:|||||| ||||||||||||:|||| |
Db         80 NLKDEVLSPSWDSLFSGSQGQLQPGARIFSFDGRDVLRHPAWPQKSVWHGSDPSGRRLME 139

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 169
              |||||||||   |||||||| |||| | |||||::||||||||||||:
Db        140 SYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMTS 188


RESULT 12
US-09-561-499-13
; Sequence 13, Application US/09561499
; Patent No. 6524583
; GENERAL INFORMATION:
;  APPLICANT:  Philip E. Thorpe
;  APPLICANT:  Rolf A. Brekken
;  TITLE OF INVENTION:  ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
;  FILE REFERENCE:  4001.002582
;  CURRENT APPLICATION NUMBER: US/09/561,499
;  CURRENT FILING DATE:  2000-04-28
;  PRIOR APPLICATION NUMBER:  60/131,432
;  PRIOR FILING DATE:  1999-04-28
;  NUMBER OF SEQ ID NOS:  44
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 191
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  SYNTHETIC
US-09-561-499-13

  Query Match             87.1%;  Score 778;  DB 4;  Length 191;
  Best Local Similarity   85.8%;  Pred. No. 5.6e-87;
  Matches  145;  Conservative   13;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||:|||||||||||||||||||||||||||:||||||||||||||||||||| :||||
Db         20 VALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRGSVPIV 79

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              |||||:| |||::|||||:| |:||||||||||:|||||| ||||||||||||:|||| |
Db         80 NLKDEVLSPSWDSLFSGSQGQLQPGARIFSFDGRDVLRHPAWPQKSVWHGSDPSGRRLME 139

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 169
              |||||||||   |||||||| |||| | |||||::||||||||||||:
Db        140 SYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMTS 188


RESULT 13
US-09-998-831-13
; Sequence 13, Application US/09998831
; Patent No. 6676941
; GENERAL INFORMATION:
;  APPLICANT: Philip E. Thorpe
;  APPLICANT:  Rolf A. Brekken
;  TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY
;  TITLE OF INVENTION:  INHIBITING VEGF
;  FILE REFERENCE: 4001.002584
;  CURRENT APPLICATION NUMBER: US/09/998,831
;  CURRENT FILING DATE:  2001-11-30
;  PRIOR APPLICATION NUMBER: 09/561,108
```

```
;  PRIOR FILING DATE: 2000-04-28
;  NUMBER OF SEQ ID NOS: 44
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 191
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  SYNTHETIC
US-09-998-831-13

  Query Match             87.1%;  Score 778;  DB 4;  Length 191;
  Best Local Similarity   85.8%;  Pred. No. 5.6e-87;
  Matches  145;  Conservative   13;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||:|||||||||||||||||||||||||:|||||||||||||||||||||| :||||
Db         20 VALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRGSVPIV 79

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              |||||:| |||::|||||:| |:||||||||||:|||||| ||||||||||||:|||| |
Db         80 NLKDEVLSPSWDSLFSGSQGQLQPGARIFSFDGRDVLRHPAWPQKSVWHGSDPSGRRLME 139

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 169
              |||||||||   |||||||| |||| | |||||::|||||||||||||:
Db        140 SYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMTS 188


RESULT 14
US-08-159-784-2
; Sequence 2, Application US/08159784
; Patent No. 5643783
;  GENERAL INFORMATION:
;    APPLICANT:  Bjorn R. Olsen
;    TITLE OF INVENTION:  NOVEL COLLAGEN AND USES THEREOF
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  MS-DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/159,784
;      FILING DATE:  December 1, 1993
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  John F. Freeman
;      REGISTRATION NUMBER:  29,066
;      REFERENCE/DOCKET NUMBER:  00246/170001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  195
;      TYPE:  amino acid
;      STRANDEDNESS:  N/A
;      TOPOLOGY:  N/A
US-08-159-784-2

  Query Match             86.8%;  Score 775;  DB 1;  Length 195;
  Best Local Similarity   85.2%;  Pred. No. 1.3e-86;
  Matches  144;  Conservative   14;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||:|||||||||||||||||||||||||:|||||||||||||||||||||| :||||
Db         24 VALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRGSVPIV 83

Qy         61 NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTE 120
              |||||:| |||::|||||:| ::|||||||||||:|||||| ||||||||||||:|||| |
Db         84 NLKDEVLSPSWDSLFSGSQGQVQPGARIFSFDGRDVLRHPAWPQKSVWHGSDPSGRRLME 143

Qy        121 SYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 169
              |||||||||   |||||||| |||| | |||||::|||||||||||||:
Db        144 SYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMTS 192


RESULT 15
US-08-985-526-36
; Sequence 36, Application US/08985526
; Patent No. 6080728
;  GENERAL INFORMATION:
;    APPLICANT:  Mixson, James A
;    TITLE OF INVENTION:  CARRIER:DNA COMPLEXES CONTAINING DNA
;    TITLE OF INVENTION:  ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
;    TITLE OF INVENTION:  THERAPY
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Connolly, Bove, Lodge, & Hutz
;      STREET:  1220 Market Street, P.O. Box 2207
;      CITY:  Wilmington
;      STATE:  Delaware
;      COUNTRY:  U.S.A.
;      ZIP:  19899
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/985,526
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/608,845
;      FILING DATE:  16-JUL-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McMorrow Jr., Robert G
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (302) 658-9141
;      TELEFAX:  (302) 658-5613
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  185 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-985-526-36

  Query Match             82.2%;  Score 734;  DB 3;  Length 185;
  Best Local Similarity   82.4%;  Pred. No. 1.3e-81;
  Matches  140;  Conservative   14;  Mismatches   14;  Indels    2;  Gaps    2;

Qy          1 VALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIV 60
              ||||:|||||||||||||||  || |||:|||||||||||||||||||||| :||||
Db         14 VALNTPLSGGMRGIRGADFQCFNNAR-VGLSGTFRAFLSSRLQDLYSIVRRADRGSVPIV 72

Qy         61 -NLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLT 119
               ||:||:| |||::|||||:| |:||||||||||:|||||| |||:|||||||:||||
Db         73 QNLRDEVLSPSWDSLFSGSQGQLQPGARIFSFDGRDVLRHPAWPQRSVWHGSDPSGRRLM 132

Qy        120 ESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 169
              |||||||||   |||||||| |||| | ||||| :|||||||||||||:
Db        133 ESYCETWRTETTGATGQASSLLSGRLLEQRAASCHDSYIVLCIENSFMTS 182
```

Search completed: March 13, 2004, 08:19:01
Job time : 23 secs